



सत्यमेव जयते

INDIAN AGRICULTURAL
RESEARCH INSTITUTE, NEW DELHI

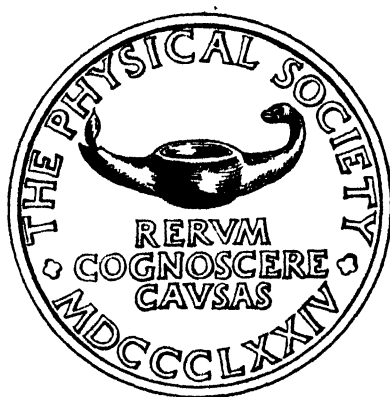
I.A.R.I.6.

GIP NLK—H.3 I.A.R.I.—10-5-55—15,000

THE
PROCEEDINGS
OF
THE PHYSICAL SOCIETY

FROM JULY 1948 TO DECEMBER 1948

VOLUME 61



Published by
THE PHYSICAL SOCIETY
1 Lowther Gardens, Prince Consort Road,
London S.W. 7

OFFICERS AND COUNCIL, 1948-49

PRESIDENT

G. I. FINCH, M.B.E., D.Sc., F.R.S.

VICE-PRESIDENTS

who have filled the Office of President

C. H. LEES, D.Sc., F.R.S.

Sir FRANK SMITH, G.C.B., G.B.E., D.Sc., F.R.S.

Sir OWEN RICHARDSON, M.A., D.Sc., F.R.S.

W. H. ECCLES, D.Sc., F.R.S.

A. O. RANKINE, O.B.E., D.Sc., F.R.S.

T. SMITH, M.A., F.R.S.

ALLAN FERGUSON, M.A., D.Sc.

Sir CHARLES DARWIN, K.B.E., M.C., M.A., Sc.D., F.R.S.

E. N. da C. ANDRADE, Ph.D., D.Sc., F.R.S.

D. BRUNT, M.A., Sc.D., F.R.S.

VICE-PRESIDENTS

A. J. PHILPOT, C.B.E., M.A., B.Sc.

W. JEVONS, D.Sc., Ph.D.

W. D. WRIGHT, D.Sc.

W. S. STILES, O.B.E., D.Sc., Ph.D.

HONORARY SECRETARIES

C. G. WYNNE, B.A. (*Business*)

H. H. HOPKINS, Ph.D. (*Papers*)

HONORARY FOREIGN SECRETARY

E. N. da C. ANDRADE, Ph.D., D.Sc., F.R.S.

HONORARY TREASURER

H. SHAW, D.Sc.

HONORARY LIBRARIAN

R. W. B. PEARSE, D.Sc., Ph.D.

ORDINARY MEMBERS OF COUNCIL

C. H. COLLIE, M.A., B.Sc.

L. F. BATES, D.Sc., Ph.D.

D. ROAF, M.A., D.Phil.

R. C. EVANS, M.A., Ph.D.

A. C. G. MENZIES, M.A., D.Sc.

L. C. MARTIN, D.Sc.

R. PEIERLS, C.B.E., M.A., Dr.Phil.,
D.Sc., F.R.S.

C. E. WYNN-WILLIAMS, D.Sc., Ph.D.

F. C. TOY, C.B.E., D.Sc.

A. G. QUARRELL, D.Sc., Ph.D.

J. H. AWBERY, M.A., B.Sc.

A. B. WOOD, O.B.E., D.Sc.

COLOUR GROUP

Chairman

J. G. HOLMES, B.Sc.

Honorary Secretary

R. G. HORNER, B.A.

LOW-TEMPERATURE GROUP

Chairman

Sir CHARLES DARWIN, K.B.E., M.C.,
M.A., Sc.D., F.R.S.

Honorary Secretary

G. G. HASELDEN, Ph.D.

OPTICAL GROUP

Chairman

L. C. MARTIN, D.Sc.

Honorary Secretary

E. W. H. SELWYN, B.Sc.

ACOUSTICS GROUP

Chairman

H. L. KIRKE, C.B.E., M.I.E.E.

Honorary Secretaries

W. A. ALLEN, B.Arch., A.R.I.B.A.

A. T. PICKLES, O.B.E., M.A.

SECRETARY-EDITOR

Miss A. C. STICKLAND, Ph.D.

1 Lowther Gardens, Prince Consort Road, London S.W.7

(Telephone : KENSington 0048)

CONTENTS

Part 1. 1 July 1948

	PAGE
Dr. G. L. PICKARD and Prof. F. E. SIMON. The atomic heats of palladium, sodium and mercury at low temperatures	1
Mr. R. B. DINGLE. The theory of the propagation of second sound in helium II	9
Prof. C. A. COULSON, Mr. D. P. CRAIG and Mr. A. MACCOLL. Electronic structure of fulvene	22
Prof. E. GLUECKAUF. Some observations concerning the energy of nuclei	25
Prof. A. DUPIER. The temperature effect on cosmic-ray intensity and the height of meson formation	34
Mr. R. I. B. COOPER. The attenuation of ultra-high-frequency electromagnetic radiation by rocks	40
Dr. G. G. MACFARLANE. The application of a variational method to the calculation of radio wave propagation curves for an arbitrary refractive index profile in the atmosphere	48
Mr. W. L. PRICE. Radio shadow effects produced in the atmosphere by inversions	59
Mr. J. C. JAEGER. Equivalent path and absorption for oblique incidence on a curved ionospheric region	78
Dr. A. TAYLOR. On a new type of rotating anode x-ray tube	86
Letters to the Editor :	
Prof. S. CHAPMAN. The dipole moment of the supposed fundamental magnetic field rotation	95
Prof. A. R. UBBELOHDE. The freezing-in of nuclear equilibrium	96
Dr. C. DODD. The detection of heating in liquid dielectrics at low frequencies	97
Dr. R. F. BARROW and Mr. M. F. R. MULCAHY. The B-N band-system of silver iodide	99
Discussion on paper by Mr. R. WEIL entitled " The variation with temperature of metallic reflectivity " (<i>Proc. Phys. Soc.</i> , 1948, 60, 8)	101
Discussion on paper by Dr. G. F. J. GARLICK and Mr. A. F. GIBSON entitled " The electron trap mechanism of luminescence in sulphide and silicate phosphors " (<i>Proc. Phys. Soc.</i> , 1948, 60, 574)	101
Reviews of books	102

Part 2. 1 August 1948

Dr. D. R. DAVIES. Shock waves in air at very high pressures	105
Mr. STEWART PATERSON. The reflection of a plane shock wave at a gaseous interface	119
Dr. E. EASTWOOD and Mr. K. A. MERCER. A study of transient radar echoes from the ionosphere	122
Dr. L. G. H. HUXLEY, Mr. H. G. FOSTER and Mr. C. C. NEWTON. Measurements of the interaction of radio waves in the ionosphere	134
Mr. R. KINGSLAKE. The diffraction structure of the elementary coma image	147
Prof. F. ZERNIKE. Diffraction and optical image formation	158
Mr. A. J. ELLEMAN and Dr. H. WILMAN. The structure and growth of PbS deposits on rocksalt substrates	164
Mr. C. M. G. LATTES, Mr. G. P. S. OCCHIALINI and Dr. C. F. POWELL. A determination of the ratio of the masses of π - and μ -mesons by the method of grain-counting	173
Mr. Y. GOLDSCHMIDT-CLERMONT, Mr. D. T. KING, Mr. H. MUIRHEAD and Mr. D. M. RITSON. Determination of the masses of charged particles observed in the photographic plate	183
Mr. E. G. MICHAELIS. A modification of Benoît's method of exact fractions	194

Letters to the Editor :

Dr. K. F. CHACKETT and Mr. G. R. MARTIN. An isotopic abundance rule	197
Miss SYLVIA A. E. HIRD and Mr. H. F. STEPHENSON. Distribution coefficients for the calculation of colours in the C.I.E. trichromatic system for a total radiator at 2450° K.	198
Reviews of books	200

Part 3. 1 September 1948

Dr. C. L. SMITH. A theory of transient creep in metals	201
Dr. C. S. CHOW. The thermal conductivity of some insulating materials at low temperatures	206
Dr. R. H. SLOANE and Mr. C. S. WATT. The emission of negative ions from oxide coated cathodes	217
Mr. D. M. MACKAY. Some new aspects of high-speed oscillography	235
Mr. W. WALKINSHAW. Theoretical design of linear accelerator for electrons	246
Mr. R. B. R.-SHERSBY-HARVIE. Travelling wave linear accelerators	255
Mr. L. B. MULLETT and Mr. B. G. LOACH. Experimental work on corrugated waveguides and associated components for linear electron accelerators	271
Mr. F. K. GOWARD. Effect of azimuthal inhomogeneities in the magnetic field of a betatron or synchrotron	284
Letters to the Editor :	
Dr. H. C. BOLTON. The measurement of dielectric constant in a waveguide	294
Reviews of books	294
Royal Society Mathematical Tables Committee	296
Corrigenda	296

Part 4. 1 October 1948

Dr. W. H. RAMSEY. On Schwinger's theory of nuclear forces	297
Mr. D. E. BUNYAN, Dr. A. LUNDBY, Mr. A. H. WARD and Dr. D. WALKER. The delayed coincidence method in the study of radioactivity, with application to isomerism in ^{181}Ta	300
Dr. G. D. ROCHESTER and Dr. C. C. BUTLER. The penetrating particles in cosmic-ray showers : I. Heavily-ionizing particles	307
Mr. J. L. PUTMAN. Analysis of spurious counts in Geiger counters	312
Dr. G. O. JONES. The mechanism of the reversible thermal after-effects in glass at ordinary temperatures	320
Dr. J. D. CRAGGS and Prof. J. M. MEEK. The initiation of low pressure glow discharges	327
Mr. A. R. BRYANT. The pressures exerted by an underwater explosion bubble on a circular target plate in a disc-like baffle	341
Dr. E. G. RICHARDSON. The impact of a solid on a liquid surface	352
Mr. J. E. CAFFYN. A correction in viscometry due to varying hydrostatic head	367
Mr. S. K. RUNCORN. The radial variation of the earth's magnetic field. With an appendix by Prof. S. CHAPMAN	373
Dr. A. R. STOKES. A numerical Fourier-analysis method for the correction of widths and shapes of lines on x-ray powder photographs	382
Letters to the Editor :	
Mr. R. STREET and Mr. J. C. WOOLLEY. A note on the ΔE effect in alnico	391

Part 5. 1 November 1948

Prof. YVES ROCARD. Sur les conditions d'auto-oscillation des systèmes vibrants	393
Prof. Sir GEORGE THOMSON. The growth of crystals. (Thirty-second Guthrie Lecture)	403
Dr. H. WILMAN. The interpretation and application of electron-diffraction "Kikuchi-line" patterns : Part II. The methods of indexing the patterns	416
Dr. J. SKREBOWSKI. Electron diffraction from anthracene with special reference to the diffuse background	430
Dr. C. GURNEY and Mr. Z. BORYSOWSKI. Delayed fracture of materials under tension, torsion and compression	446
Prof. N. FEATHER and Dr. H. O. W. RICHARDSON. β -disintegration and the Sargent diagram	452

Contents

v
PAGE

Prof. N. FEATHER, Mr. J. KYLES and Dr. R. W. PRINGLE. Investigations using a permanent magnet double β -ray spectrograph with coincidence counting :	
I. Experiments with thorium active deposit	466
Mr. S. BYARD. Note on the impedance variations of an electro-acoustic transducer in a reflecting field	478
Letters to the Editor :	
Dr. C. STRACHAN. Dependence of β -decay on the sign of the change in nuclear spin	481
Dr. J. S. MCPETRIE and Dr. J. A. SAXTON : Mr. R. I. B. COOPER. The attenuation of ultra-high frequency electromagnetic radiation by rocks	482
Mr. P. C. THONEMANN, Mr. J. MOFFATT, Dr. D. ROAF and Mr. J. H. SANDERS. The performance of a new radio-frequency ion source	483
Mr. C. H. COLLIE and Mr. M. A. GRACE. The absorption coefficient of the β rays from ^{35}S in aluminium	485
Mr. O. SIMPSON. The effect of room temperature radiation on the infra-red response of lead telluride photoconductors	486
Reviews of books	488

Part 6. 1 December 1948

Dr. MARY P. LORD. The measurement of the eye movements of a prone observer during monocular fixation	489
Dr. E. WOLF. On the designing of aspheric surfaces	494
Mr. L. C. TSIEN. The interference pattern of two neighbouring radiations in a Hilger prism interferometer	504
Mr. A. J. KENNEDY. A method of fitting the Andrade creep equation to experimental results	510
Dr. LARS BECKMAN. The influence of a transverse magnetic field on a cylindrical plasma	515
Dr. S. WAGENER. Note on the ionic conductivity of oxide-coated cathodes	521
Dr. R. M. CHAUDHRI and Mr. A. W. KHAN. Emission of secondary electrons from nickel and molybdenum by neutral atoms of mercury and potassium	526
Dr. F. C. CHAMPION and Dr. R. R. ROY. The scattering of fast β -particles through large angles by nitrogen nuclei	532
Dr. G. D. ROCHESTER and Dr. C. C. BUTLER. The penetrating particles in cosmic-ray showers : II. The lightly-ionizing penetrating particles in penetrating showers	535
Dr. C. J. GORTER. A few remarks about spectroscopy at low frequencies	541
Mr. D. M. S. BAGGULEY, Dr. B. BLEANEY, Dr. J. H. E. GRIFFITHS, Dr. R. P. PENROSE and Miss B. I. PLUMPTON. Paramagnetic resonance in salts of the iron group—A preliminary survey : I. Theoretical discussion	542
Mr. D. M. S. BAGGULEY, Dr. B. BLEANEY, Dr. J. H. E. GRIFFITHS, Dr. R. P. PENROSE and Miss B. I. PLUMPTON. Paramagnetic resonance in salts of the iron group—A preliminary survey : II. Experimental results	551
Dr. H. R. I. LAMONT. Atmospheric absorption of millimetre waves	562
Dr. E. E. SCHNEIDER. Note on nuclear magnetic resonance	569
Corrigenda	571
Abstracts of papers read at Oxford Conference	572
Letters to the Editor :	
Mr. P. JENSEN. Nuclear photo-effect and isotopic number	579
Mr. F. K. GOWARD and Mr. J. J. WILKINS. Validity of two-dimensional design of synchrotron pole-faces	580
Reviews of books	582
Obituary notices :	
Dr. ALEXANDER MÜLLER	585
Sir CLIFFORD COPLAND PATERSON, O.B.E., D.Sc., F.R.S.	587
Emeritus Professor S. W. J. SMITH, M.A., D.Sc., F.Inst.P., F.R.S.	591
Index to Volume 61	594
Index to Reviews of books, Vol. 61	600
Proceedings at the meetings, 1947-48	vi
Report of Council for the year 1947	xvi
Report of the Honorary Treasurer for the year 1947	xxi

PROCEEDINGS AT THE MEETINGS OF THE PHYSICAL SOCIETY

SESSION 1947-48

24 September 1947

Extraordinary General Meeting, at Imperial College, London S.W.7. The President, Professor G. I. Finch, was in the Chair.

The following Special Resolution was passed *nem. con.* :

That the Articles of the Association of the Society be altered as follows :—

- (i) Article 11, by substituting
 - (a) in line 1, for the word “ two ” the word “ three ”;
 - (b) in line 4, for the words “ one guinea a year ” the words “ two guineas a year, in accordance with Article 31 ”.
- (ii) Article 31, by substituting
 - (a) in line 2, for the words “ two guineas or one guinea ” the words “ three guineas or two guineas ”; and
 - (b) in lines 11 and 22, for the words “ two guineas or may pay an annual subscription of one guinea ” the words “ three guineas or may pay an annual subscription of two guineas ”.
- (iii) Article 42, by substituting for the words “ ten shillings and sixpence ” the words “ fifteen shillings ”.

24 September 1947

Science Meeting, at Imperial College, London S.W.7. The President, Professor G. I. Finch, was in the Chair.

It was announced that Council had elected the following to Student Membership : John Robert Balbin, Benjamin Robert Malcolm, James Ruddick.

The following were elected to Fellowship : Charles Fountain Brockelsby, Dorothy Gertrude Fisher, Rt. Hon. the Earl of Halsbury, Gilbert Henry Hinds, James Lawrence King, Roy Flavell King, Edward Lee, Francis Edmund Jury Ockenden, Kenneth Charles Worton Pedder, Edward James William Underhill.

The fifteenth Thomas Young Oration was delivered by Professor Fritz Zernike (Groningen), whose subject was “ Diffraction and optical image formation ”.

1 October 1947

The thirty-fourth meeting of THE COLOUR GROUP, at Imperial College, London S.W.7. Mr. J. G. Holmes was in the Chair.

A discussion was held of the International Colour-Vision Conference held at Cambridge.

3 October 1947

The twenty-ninth meeting of THE OPTICAL GROUP at Imperial College, London S.W.7. Professor L. C. Martin was in the Chair.

The following papers were read and discussed :

- “ The optical sine condition and Herschel's condition ”, by H. H. Hopkins;
- “ Eikonal theory and aspheric systems ”, by E. Wolf;
- “ On the disturbance near the focus of waves of radially non-uniform amplitude ”, by H. H. Hopkins.

9 October 1947

The second meeting of THE ACOUSTICS GROUP, at Trinity College of Music, London W.1.
Mr. H. L. Kirke was in the Chair.

A discussion was held on "The place of the formant in the production of musical tone".

31 October 1947

Science Meeting, at Imperial College, London S.W.7. The President, Professor G. I. Finch, was in the Chair.

It was announced that Council had elected the following to Student Membership : Keith Homer Barker, Joan Reith Bennett, Peter Jesse Bennett, Geoffrey Waddington Benson, Francis Geoffrey Blackler, Derek Bloomer, Victor Braxton, William Brodie, Joyce Esmé Brooks, John Robert Billson Calow, John Carruthers, Peter R. Cuthbert, Kentish B. Dawson, Joan Edwards, Tony Frederick Wallace Embleton, Henry Lewis Evans, Andrzej Folkierski, Peter Henry Harris, Peter Heggs, Anthony Leonard Hendon, Albert J. Herz, Desmond Peter Howlett, Peter Denis Lee, Francis John Lowes, Eileen Eleanor Mallett, Basil John Mason, Norman Morris, Frank Donald Morten, Raymond Geoffrey Todd Munday, Keith Roland Prosser, Brian William Purslow, Malcolm John Ravenscroft, Vera Winifred Reed, Vernon Timbrell, Peter Bernard Watt, David Edward Weston, John Willis.

The following were elected to Fellowship : Francis Cecil Chalklin, William Marshall Collier, James Brownlee Dick, Philip Gambrell Gibbs, Geoffrey Bond Harrison, John Lee Howarth, Francis Edgar Jones, Martin Ludwig Loehr (transferred from Student Membership), Ewen M'Ewen, Peter Crompton Russell, David Shoenberg, Alan Alfred Ware, James Rennie Whitehead, Geoffrey Musk Wilcox.

The following papers were read and discussed :

"Diffusion pumps : a critical discussion of existing theories", by D. G. Avery and R. Witty;

"Electron diffraction from anthracene, with special reference to the diffuse background", by J. Skrebowski.

8 November 1947

Science Meeting, at University College, Nottingham. The President, Professor G. I. Finch, was in the Chair.

The following papers were read and discussed :

"The adiabatic temperature changes accompanying magnetization in low and moderate fields", by L. F. Bates and E. G. Harrison;

"The Barkhausen effect", by H. D. Bush and R. S. Tebble;

"The variation with magnetization of Young's modulus for cobalt", by R. E. Street;

"Thermionic emission and band overlap", by E. P. Wohlfarth.

The following demonstrations were given :

"The magnetic sorting bridge", by N. Underwood;

"Bitter figures", by F. E. Neale.

13 November 1947

The third meeting of THE ACOUSTICS GROUP, at Imperial College, London S.W.7.
Mr. H. L. Kirke was in the Chair.

The following paper was read and discussed :

"Supersonic absorption of liquids", by J. M. M. Pinkerton.

21 November 1947

Science Meeting, at the Science Museum, London S.W.7. The President, Professor G. I. Finch, was in the Chair.

It was announced that the Council had elected the following to Student Membership : John Jones Benbow, Noel Keith Bridge, Llewellyn John Elliott, Marie Ellis, Rita Barbara

Joan Francis, Reginald Leslie Heasman, Nora Evelyn Hill, Geoffrey Morgan David Bruce Jones, Eleanor Marian Leaf, Joyce Dorothy Ling, K. H. Marlow, Philip Rhodes, Barbara Ethel Sabey, John Albert Charles Samms, Dennis Frederick Shaw, Eric Thomas Smith, David Henry Trevena, Edward Victor Vernon, Stuart Astley Young.

The following were elected to Fellowship : Samuel Baxter, Israel Berz, Albert John Davies, Arthur Desmond Dobson, Ali Abd Elkerim Elrahim, Christopher Leonard Sidney Gilford, Heiz Kurt Henisch, Cornelius Graham Mayer, John Henry Mitchell, Edmund Richard Drake Seager, Narayan Chandra Sen Gupta, Harold Neville Vazeille Temperley, Joseph West.

The thirty-first Guthrie Lecture was delivered by Professor J. D. Bernal (Birkbeck College, London), whose subject was "The physical basis of life".

26 November 1947

The thirtieth meeting of THE OPTICAL GROUP, at Northampton Polytechnic, London E.C.1. Professor L. C. Martin was in the Chair.

The following papers were read and discussed :

"Quantum fluctuations and visual acuity at low brightnesses", by M. H. Pirenne;

"Experiments on flash visual acuity", by D. J. Day;

"The measurement of the eye movements of a prone observer during monocular fixation", by Mary P. Lord.

3 December 1947

The thirty-fifth meeting of THE COLOUR GROUP, at the Geological Museum, S.W.7. Mr. J. G. Holmes was in the Chair.

The following paper was read and discussed :

"Ultramarine—natural and synthetic", by J. Pryce-Jones.

4 December 1947

Science Meeting, a joint meeting with the Institution of Electrical Engineers, at the Institution of Electrical Engineers, London W.C.2. Mr. P. Good was in the Chair.

The following paper was read and discussed :

"Speech communication under conditions of deafness or of loud noise", by W. G. Radley.

4 December 1947

The fourth meeting of THE ACOUSTICS GROUP, at the Institution of Electrical Engineers, London W.C.2, held as a joint meeting of the Physical Society and the Institution of Electrical Engineers. (See above)

12 December 1947

Joint meeting of THE OPTICAL GROUP with the Industrial Spectroscopic Group of the Institute of Physics, at Imperial College. Mr. D. M. Smith was in the Chair.

The following demonstrations were given :

"Microphotometer using cathode-ray oscilloscope", by R. Fürth and W. D. Oliphant.

"Improvements to non-recording microphotometers by Adam Hilger Limited", by A. C. Menzies.

The following paper was read and discussed :

"The contribution of optical scatter to the systematic error in the microphotometry of spectral lines", by D. R. Barber.

18 December 1947

The eleventh meeting of THE LOW TEMPERATURE GROUP, at the Science Museum, London S.W.7. Sir Alfred Egerton was in the Chair.

A discussion entitled "Insulating materials and methods for use at low temperatures" was opened by F. E. Simon, who was followed by Ezer Griffiths, A. M. Clark and Sir Alfred Egerton.

19 December 1947

Science Meeting, at the Science Museum, London S.W.7. The President, Professor G. I. Finch, was in the Chair.

It was announced that the Council had elected the following to Student Membership : Geoffrey Martin Brown, Ian Clifford Cheeseman, John Crawford Fiddes, Ann Pauline Korndorffer, Robert Barclay Nicholls, Eric Griffith Samuel, Edgar Albert George Shaw.

The following were elected to Fellowship : Ralph Leonard Aspen, Dirk Bijl, Douglas Richard Chick, Merrill Francis Distad, Brian Donovan, Clifton Darfield Ellyett, Albert Sydney Freeborn, Harold Hall Keen, Dennis John Tarran (transferred from Student Membership).

The fourth Charles Chree Address was delivered by Sir Edward Appleton, whose subject was "Geomagnetism and the ionosphere".

The third Charles Vernon Boys Prize Address was delivered by Dr. C. F. Powell (Bristol), whose subject was his work on the development of the photographic plate technique in the investigation of fundamental particles.

8 January 1948

The fifth meeting of THE ACOUSTICS GROUP, at the National Hospital, London W.C.1. Mr. H. L. Kirke was in the Chair.

The following papers were read and discussed :

"The use of supersonic sound by animals", by H. Hartridge;

"Tone deafness", by D. Fry;

A film and demonstration, "Pure-tone audiometry in young children : a new technique", were given by C. S. Hallpike and M. R. Dix.

9 January 1948

Science Meeting, at Imperial College, London S.W.7. The President, Professor G. I. Finch, was in the Chair.

It was announced that the Council had elected the following to Student Membership : Max Michael Bluhm, Geoffrey William Dolphin, Manapragada Venkata Sree Ramanchandra Murty, Natesan Ramanathan.

The following were elected to Fellowship, the last seven being transferred from Student Membership : George Emile Fishter, John Barrett Hasted, Elsie Anna Grace Knowles, Douglas Clement Spanner, Carl Henry Westcott, John Charles Willmott, Hyman Woolfe; Donald Kenneth Ashpole, Robert Guy Chambers, John Francis Irving Cole, Michael William Feast, Peter Henry Rose, Vera Mary Scamans, Ian Murray Watt.

The following papers were read and discussed :

"Shock waves in air at very high pressures", by D. R. Davies;

"The propagation of shock waves in steel and lead", by D. C. Pack, W. M. Evans and H. J. James.

28 January 1948

The thirty-sixth meeting of THE COLOUR GROUP, at the National Gallery, London W.C.2.

A discussion was held on the Cleaned Pictures Exhibition at the National Gallery.

Proceedings at meetings

6 February 1948

Science Meeting, at Imperial College, London S.W.7. The President, Professor G. I. Finch, was in the Chair.

It was announced that the Council had elected the following to Student Membership : William Bosley, Kenneth Charles Ellis, John Ewles, Hiroshi Julian Goldsmid, Albert Leslie Hodson, Thomas Oliver Jeffries, Arthur Eric Jenkinson, Kenneth Herbert Jolliffe, Pallavar Neelakantan Krishnamoorthy, John Smith Mair, Geoffrey Edward Perry, Ruth Helen Rogers, Charmian Cecil Ruth Sinclair, David Smith, Geoffrey Stephenson, John Weston Talbot, Guy Kendal White.

The following were elected to Fellowship, the last eleven being transferred from Student Membership : Francis Kenneth Earp, Jack Ewles, Leslie Arthur Holland, Leslie Knopp, Leonard Alfred Lacey, Anthony Spence Merriweather; Deryck Russell Bell, Antony Harley Boston, James Henry Ellis, Stanley Gill, Norman Edward Goddard, Ronald Henry Hall, Denis Noel Layton, Peter Albert Michaels, David Edwin Umfreville Ridout, Colin Henry Smith, Guy Charles Tavernier.

The following papers were read and discussed :

“ The efficiency of counting systems ”, by M. Blackman and J. L. Michiels;

“ The moving-coil galvanometer ”, by N. F. Astbury.

11 February 1948

The thirty-seventh meeting of THE COLOUR GROUP, at Imperial College, London S.W.7. Mr. J. G. Holmes was in the Chair.

The following paper was read and discussed :

“ The classification of colour mixture phenomena ”, by D. R. Duncan.

12 February 1948

The sixth meeting of THE ACOUSTICS GROUP, at the Science Museum, London S.W.7. Mr. H. L. Kirke was in the Chair.

A discussion was held on “ The borderline of psychology, physics and music ”.

13 February 1948

The twelfth meeting of THE LOW TEMPERATURE GROUP. A Joint Conference with the Institution of Chemical Engineers, entitled “ Developments in the Industrial Production and Use of Gaseous Oxygen ”, at the Institution of Civil Engineers, Great George Street, London S.W.1.

Morning Session. Sir Charles Darwin was in the Chair.

The following papers were read and discussed :

“ The use of oxygen for metallurgical heating in the iron and steel industry ”, by M. W. Thring;

“ The use of oxygen for town gas manufacture ”, by F. J. Dent;

“ The oxygen diagram ”, by S. K. Bhattacharrya and D. M. Newitt.

Afternoon Session. Mr. H. W. Cremer was in the Chair.

The following papers were read and discussed :

“ Theoretical aspects of air separation ”, by M. Ruhemann;

“ Modern gaseous oxygen production methods ”, by P. M. Schuftan;

“ Some notes on the economics of medium purity oxygen ”, by J. R. Park.

11 March 1948

The seventh meeting of THE ACOUSTICS GROUP, at the Royal Society of Arts, London W.C.2. Mr. H. L. Kirke was in the Chair.

A discussion was held on “ Loudspeakers, with special reference to high-fidelity monitoring reproducers ”.

12 March 1948

Science Meeting, at Imperial College, London S.W.7. The Honorary Business Secretary, Professor A. O. Rankine, was in the Chair.

It was announced that the Council had elected the following to Student Membership : Richard James Courshee, John Keith Donoghue, John Alfred Gurdler, Arthur John Heard, Albert George Hill, John Noel Hodgson, Valerie Jackson, John Compton Lawman, Geoffrey Charles Edward Olds, Jorgen Lykke Olsen, Ernst Rabinowicz, Cyril Scott, Guy Frederic Arthur de Visme; George Herbert Avery Cole, Antony Julian Croft, Stella Mayne, Guy Edward Williamson-Noble.

The following were elected to Fellowship; the last seven being transferred from Student Membership : Andre Jean Louis Bertholot, Francis Donald Boardman, Bertram Vivian Bowden, Clifford Wallace Collins, William Ronald Myers, John Robert Pikett, John Charles Simmonds, Beryl Freda Thorne, James Leslie Tuck, Putcha Venkateswarlu, Derek John Webb; Robert Bernard Clark, Robert Lester Cowley, Ronald Bertie Ide, Donald Hill Perkins, Derek Malcolm Thomas, Richard John Wakelin, Peter Robert Wyke.

The following papers were read and discussed :

"Electron trap mechanism of luminescence in sulphide and silicate phosphors", by G. F. J. Garlick and A. F. Gibson;

"Magnetic dispersion at microwave frequencies : I. The measurement of high frequency permeability", by G. Eichholz and G. F. Hodsman.

19 March 1948

The thirty-eighth meeting of THE COLOUR GROUP, at Imperial College, London S.W.7. Dr. W. S. Stiles was in the Chair.

The following paper was read and discussed :

"The physical structure of coloured glass", by J. G. Holmes.

The meeting was preceded by the eighth *Annual General Meeting of the Colour Group* for the presentation of the Committee's report on the work of the Group in 1947-48, the election of Officers and Committee for 1948-49 and the presentation of the Chairman's Address on "The aims and activities of the Colour Group".

2 April 1948

The thirty-first meeting of THE OPTICAL GROUP, at Imperial College, London S.W.7. Mr. B. K. Johnson was in the Chair.

The following paper was read :

"Wide aperture mirror lens systems", by C. G. Wynne.

The following demonstrations were given :

"Method of keeping the brightness of a lamp constant without using a voltmeter", by R. W. G. Hunt

"Mechanism for obtaining a linear density scale with polarizing light modulators", by R. W. G. Hunt.

5 April 1948

The thirteenth meeting of THE LOW TEMPERATURE GROUP, at the Chemical Society, Burlington House, Piccadilly, London W.1. First Joint Meeting with the Institute of Refrigeration. Sir Charles Darwin was in the Chair.

The following paper was read and discussed :

"Heat pump problems and industrial realizations", by G. Eichelberg.

Proceedings at meetings

8 April 1948

Annual General Meeting of THE ACOUSTICS GROUP, at the Royal Institute of British Architects, London W.1. Mr. H. L. Kirke was in the Chair.

The minutes of the Annual General Meeting held on 19 February 1947, and the General Meeting held on 26 June 1947, were read and confirmed.

The Officers and Committee for 1948-49 were elected.

Votes of thanks were accorded to The Physical Society for the help given during the year and to the Royal Institute of British Architects for the excellent accommodation at meetings.

8 April 1948

The eighth meeting of THE ACOUSTICS GROUP, at the Royal Institute of British Architects, London W.1. Mr. H. L. Kirke was in the Chair.

A discussion was held on "The use of microphones and loudspeakers in buildings and sound amplification and distribution", opened by H. Bagenal.

28 April 1948

The thirty-ninth meeting of THE COLOUR GROUP, at the Royal Photographic Society, Prince's Gate, London S.W.7. Mr. J. G. Holmes was in the Chair

A discussion was held on the Report on Colour Terminology, and the British proposals to be submitted to the Commission Internationale de l'Eclairage were reported.

30 April 1948

The thirty-second meeting of THE OPTICAL GROUP, at Imperial College, London S.W.7. Professor L. C. Martin was in the Chair.

The following paper was read and discussed :

"Some further applications of multiple beam interferometry", by S. Tolansky.

5 May 1948

Annual General Meeting at the Royal Institution, Albemarle Street, London W. 1. The President, Professor G. I. Finch, was in the Chair.

The minutes of the Annual General Meeting held on 8 May 1947 were read and confirmed.

The Reports of the Council and the Honorary Treasurer and the Annual Accounts for 1947 were adopted.

The Officers and Council and the Auditors for 1948-49 were elected.

Votes of thanks were accorded to the Rector and Governing Body of Imperial College and Sir George Thomson; the Director of the Science Museum; the Managers of the Royal Institution; the Royal Commission for the Exhibition of 1851; the Officers and Members of the Committees of the Colour, Low Temperature, Optical and Acoustics Groups; to the Secretaries and to the retiring Officers and Council.

The meeting was followed by an *Extraordinary General Meeting* for the election of Professor Ernest Orlando Lawrence to Honorary Fellowship of the Society.

5 May 1948

Science Meeting, at the Royal Institution, Albermarle Street, London W.1. The President, Professor G. I. Finch, was in the Chair.

It was announced that Council had elected the following to Student Membership : Robert Owen Davies, Derek Roy Holmes, Leslie George Leech, William Norman Mair, Jack Gordon Powles.

The following were elected to Fellowship, the last five being transferred from Student Membership : David Kevin Bewley, Kenneth Ellis, Alfred Graham, Thomas Burton

Grimley, William Henry Hill, John Aylward James, Samuel Lattimore, Alen De la Mare Le Claire, George Francis Longster, Arthur James Potts, Antonio Rostagni, Harold Michael Scott-Smith, Roy Samuel Sharpe, Kenneth Jack Standley, Thomas Lewis Tippell, Etienne Joseph Vassy, James Swanton Wolfe; Peter Crosby, Geoffrey Gunther Eichholz, David Rowland McCall, Peter Anthony Tanner, William Harris Thorning.

The third Holweck Discourse was delivered by Professor Y. Rocard, whose subject was "Sur les conditions d'auto-oscillation des systèmes vibrants".

4 June 1948

Science Meeting, at the Science Museum, London S.W.7. The President, Professor G. I. Finch, was in the Chair.

It was announced that the Council had elected the following to Student Membership: Gerald James Ames, Richard Edward Copland, Frank Rhoden, Patrick Walsh Seddon.

The following were elected to Fellowship, the last three being transferred from Student Membership: Oliver Martin Ashford, Frank Joseph Crow, Edgar Frank Daly, Leonard John Davies, John Hancock Davis, Allan John Ede, De Larue Everad Edmunds, John McKean Ellison, K. Abraham George, Hans H. Halban, Hans Walter Loeb, Stanislav Loria, N. K. Mundle, Asger Kjerbye Nielsen, Arthur C. Parker, John Philip Manning Prentice, Amherst Felix Thomson, Arthur Eric Jarvis Vickers; Derek Young Coomber, Peter Landsberg, Henry Thomas Richardson.

The thirty-second Guthrie Lecture was delivered by Sir George Thomson, whose subject was "The growth of crystals".

9 June 1948

The fortieth meeting of THE COLOUR GROUP, at the Lighting Service Bureau, 2 Savoy Hill, London W.C.2. Mr. J. G. Holmes was in the Chair.

The following paper was read and discussed:

"The effects of daylight and tungsten light adaptation on colour perception", by R. W. G. Hunt.

11 June 1948

The thirty-third meeting of THE OPTICAL GROUP, at Imperial College. Professor L. C. Martin was in the Chair for the first part of the meeting and Mr. T. Smith for the second.

The following papers were read and discussed:

"The computation of large numerical aperture telescope objectives", by T. Smith;

"The boundary wave theory of image formation", by L. C. Martin.

14 June 1948

The fourteenth meeting of THE LOW TEMPERATURE GROUP, at the Chemical Society, Burlington House, Piccadilly, London W.1. Sir Charles Darwin was in the Chair.

The following paper was read and discussed:

"Expansion engines for helium liquefaction and oxygen production", by S. C. Collins.

2 July 1948

Science Meeting, at Imperial College, London S.W.7. The President, Professor G. I. Finch, was in the Chair.

It was announced that Council had elected the following to Student Membership : Christopher Herbert McKoen, Nora Page, Anthony Vivian Smith de Renck, Peter Lawrence Siday.

The following were elected to Fellowship : William Jasper Clark, William Kenneth Donaldson, Charles Leslie Dyer, Sidney Walter Gogh, Wallace Russell Harper, Herbert Kitchener Hollingsworth, John Oliver Isard, Helmut Myknon, James Francis Tait, David Hardy Whiffen.

The following papers were read and discussed :

"The theory of radiation damping", by J. Hamilton;

"An extension of Kapitza's theory of delta radiation", by R. M. Sillitto.

14, 15, 16 July 1948

The ninth meeting of THE ACOUSTICS GROUP, at the Royal Institute of British Architects, London W.1. Mr. H. L. Kirke was in the Chair at the first, second and fifth sessions, Mr. H. Bagenal at the third session, and Mr. W. West at the fourth session.

The following papers were read and discussed :

"Sound transmission through partitions", by L. L. Baranek;

"Sound transmission through partitions", by G. H. Aston;

"Sound transmission through partitions", by T. Vogel;

"Absorption by elastic layers", by W. Furrer;

"Absorption by panels at oblique incidence", by L. Cremer;

"Transmission through party walls", by W. A. Allen;

"Proposals for international standards for measurement" (presented on behalf of working committee), by Fritz Ingerslev, P. H. Parkin, C. W. Kosten and G. H. Aston.

"Insulation of wood floors", by G. H. Aston;

"Impact sound", by C. W. Kosten and J. van den Eijk;

"Insulation by discontinuous structures", by W. A. Allen;

"Floating floors", by H. R. Humphreys;

"Insulation measurement for footstep noises", by P. Kipper;

"Survey of domestic noise", by W. A. Allen;

"Sound insulation in experimental dwellings at Rotterdam", by C. W. Kosten and J. van den Eijk;

"Similarity tests", by C. W. Kosten;

"Noise in schools", by J. L. Burn;

"Noise in ships", by J. Grunenwaldt;

"Relation between the various absorption coefficients", by J. Grunenwaldt;

"Noise in aeroplanes", by R. Lehmann and P. Chavasse;

"Sound insulation of flats", by P. H. Parkin and H. R. Humphreys;

"Acoustics of old theatres", by François Canac;

"Noise meters", by R. S. Dadson;

"Objective noise meters", by A. J. King;

"A mains-driven objective noise meter", by C. A. Mason;

"Objective measurement of noise", by P. Baron;

"Certain aspects of noise measurement", by D. Maurice;

"A noise analyser", by Lord Halsbury;

"Complex sounds and artificial noises for use in acoustic tests", by P. Chavasse;

"Attenuation of noise in ducts", by P. V. Bruel;

"Absorption by porous materials", by C. W. Kosten and colleagues;

"Noise level in ducts", by J. Grunenwaldt;

"Absorption", by A. Schoch;

"Noise of lighting chokes", by C. A. Mason;

"Properties of rubber", by E. Meyer;

"Measurement and analysis of machinery noise", by C. H. Bradbury.

23-24 July 1948

Science Meeting, at the Clarendon Laboratory, Oxford. The President, Professor G. I. Finch, was in the Chair for the opening session.

The following were elected to Fellowship : J. C. Jaeger, R. S. Khot, D. H. Parkinson.

The following papers were read and discussed :

" Spectroscopy at radio and radar frequencies ", by C. J. Gorter;

" The theory of line breadths in microwave and radio frequency spectra ", by J. H. Van Vleck;

" Experimental investigations of paramagnetic resonance lines ", by J. H. E. Griffiths;

" Theory of the line breadth in copper sulphate ", by M. H. L. Pryce;

" Paramagnetic resonance at low temperatures ", by B. Bleaney;

" Pressure broadening of the ammonia spectrum ", by B. Bleaney;

" Microwave absorption in gases ", by M. H. L. Pryce;

" Atmospheric absorption of millimetre waves ", by H. R. L. Lamont;

" Nuclear effects in microwave absorption in molecules ", by A. G. Hill;

" Magnetic relaxation of nuclei ", by N. Bloembergen;

" Experiments on nuclear resonance at low temperatures ", by B. V. Rollin;

" Effects of nuclear electric quadrupole moments in the R.F. spectra of solids and liquids ", by R. V. Pound;

" Measurements of relaxation times ", by L. E. Drain;

" Investigations of ferromagnetism in the microwave region ", by J. H. E. Griffiths;

" Microwave resonance absorption in ferromagnetic materials ", by C. Kittel;

" Microwave experiments with superconductors ", by A. B. Pippard;

" Vapour pressure of helium 3 ", by J. de Boer;

" The separation of helium 3 by liquid phase methods ", by J. G. Daunt;

" Experiments on helium films ", by J. B. Brown.

REPORT OF COUNCIL FOR THE YEAR ENDED 31 DECEMBER 1947

INTRODUCTORY AND GENERAL

1947 was a year of remarkable growth and activity. The increase of total membership exceeded that in any year since 1932, when the present Physical Society was formed by the amalgamation of the Optical Society and the Physical Society of London. More Science Meetings were held than in any year since 1939, and the attendances were satisfactory. A second post-war Exhibition was held with notable success in spite of the difficulties caused by the national fuel crisis and the hard winter. Two important conference reports were published. The Acoustics Group was inaugurated. These achievements were accompanied, however, by a serious financial deficit, which, though not entirely unexpected, proved to be even greater than that in 1946. Steps have been taken to bring about a considerable improvement of the financial position in 1948.

Dr. J. F. Allen resigned his membership of the Council on taking up his appointment to the Chair of Natural Philosophy at St. Andrews in September, and Dr. W. Jevons resigned the Honorary Business Secretaryship at the end of the year. The Council appointed Professor L. C. Martin and Professor A. O. Rankine, respectively, as their successors.

Dr. A. C. Stickland was appointed as Secretary-Editor, and took up her duties on 1 September.

The continued generosity of the Royal Commission for the Exhibition of 1851 in providing accommodation for the Society's headquarters at 1 Lowther Gardens is again gratefully acknowledged.

The Council again records the cordial thanks of the Society to the Rector and Governing Body of Imperial College and Professors Sir George Thomson and H. V. A. Briscoe for the great privilege of holding the Exhibition in the Physics and Inorganic Chemistry Departments of the College. For the use of Lecture Theatres for Science Meetings of the Society and its four Groups the Council thanks Sir George Thomson, the Director of the Science Museum, the Electric Lamp Manufacturers' Association, the Director of the Geological Museum, and the Councils of the Royal Photographic Society and the Royal Institute of British Architects.

MEETINGS

The Annual General Meeting was held at Imperial College, on 8 May, for the presentation and adoption of the Reports of the Council and the Treasurer and the Accounts and Balance Sheet for 1946, and for the election of the Officers and Council for 1947-48.

Three Extraordinary General Meetings were held during the year, all at Imperial College. At the first, on 28 March, a new Article 36, fixing a scale of composition fees for Life Fellowship, was adopted. At the second, on 27 June, a proposal to replace *The Proceedings of the Physical Society* by *The Magazine of Physics* was discussed but eventually withdrawn in view of opposition of Fellows. At the third, on 24 September, a special resolution to increase the basic annual subscriptions of Fellows and Student Members to three guineas and to fifteen shillings, respectively, was adopted.

Fifteen Science Meetings, in addition to those of the four specialist Groups, were held: seven in the Physics Department of Imperial College, five at the Science Museum, one at the Institution of Electrical Engineers (a joint meeting with the Institution for a discussion on speech communication under conditions of deafness and of loud noise), one in the Physics Department of the University of Manchester and one in the Physics Department of University College, Nottingham. These two provincial meetings were a continuation of the programme initiated at Birmingham in 1946, and were devoted to a discussion on the radio detection of meteors and comets and to the presentation of demonstrations and papers on magnetism, respectively.

Of the twelve ordinary meetings in London, six were given to the reading and discussion of papers, three to memorial lectures and presentations reported in later paragraphs, and one each to a discussion on the spark discharge and to a lecture-survey on multiple-beam interferometry. At the meeting on 8 May the retiring President, Professor Brunt, delivered his address on "Some aspects of the heat balance of the human body".

EXHIBITION OF SCIENTIFIC INSTRUMENTS AND APPARATUS *

The 31st Exhibition was held at Imperial College on 9–12 April. The change of date from winter to spring proved to be of great convenience to exhibitors and visitors alike. The total attendance was again high (about 8000 excluding holders of passes—members' and exhibitors'), though not so large as in 1946, and conditions were much more comfortable. The improvement was effected by the extension of the period of the Exhibition from three to four days and the controlled distribution of dated and numbered tickets for each special morning session and for each ordinary afternoon-and-evening session.

The preparation of the Catalogue was started in good time but was seriously interrupted by the national fuel crisis and electrical shut-down early in the year, and no copies were available in time for the Exhibition, but a substantial number have been sold subsequently, and others have been issued to the Export Promotion Department of the Board of Trade in view of its acknowledged value as a work of reference.

MEMORIAL LECTURES AND AWARDS**GUTHRIE LECTURE**

The 31st Guthrie Lecture was delivered at the Science Museum on 21 November by Professor J. D. Bernal, who took as his subject "The Physical Basis of Life".

THOMAS YOUNG ORATION

The 15th Oration was delivered by Professor F. Zernike at Imperial College on 24 September, the subject being "Diffraction and Optical Image Formation".

DUDDELL MEDAL

The presentation of the 23rd (1946) Duddell Medal to Dr. K. Weissenberg, the award of which was reported last year, took place at the Science Museum on 10 January; the medallist afterwards delivered a discourse on his work on x-ray goniometers.

The 24th (1947) Medal has been awarded to Dr. R. J. Van de Graaff for his invention and development of the high-voltage electrostatic generators that bear his name. It has been arranged that the presentation will be made by His Excellency the British Ambassador (Lord Inverchapel) at the Washington meeting of the American Physical Society on 29 April 1948.

CHARLES CHREE MEDAL AND PRIZE

The 4th (1947) Medal and Prize were presented at the Science Museum on 19 December to Sir Edward Appleton in recognition of his investigations of the ionosphere by radio methods. The medallist afterwards delivered a Charles Chree Address on "Geomagnetism and the Ionosphere".

CHARLES VERNON BOYS PRIZE

The 2nd (1946) Prize, the award of which was reported last year, was presented to Mr. R. W. Sutton at the Science Museum on 10 January.

The Council made the 3rd (1947) award of the Prize to Dr. C. F. Powell for his development of the photographic-plate technique for the investigation of high-speed fundamental particles, and the presentation took place at the Science Museum on 19 December.

Each of the prizemen delivered a discourse on his work.

HOLWECK PRIZE AND MEDAL

The second (1947) award of the Holweck Prize of the Physical Society and Holweck Medal of the Société Française de Physique was made to Professor E. N. da C. Andrade in recognition of his investigations of the properties of the solid and liquid states. The presentation took place at a meeting of the Société in Paris on 24 May, when Professor Andrade delivered a Holweck Discourse on "L'Ecolement des Métaux".

At his request the prize has been devoted to the purchase of books for the library of the Société Française de Physique.

PUBLICATIONS

PROCEEDINGS

Volume 59 (1947) appeared, as usual, in six two-monthly parts. Both the size of the volume and the number of copies printed were greater than for the preceding volume ; this was made possible by an increase in the official allowance of paper from 100% to 144% of the amount used in the twelve months immediately before the war. The space occupied by advertisements was slightly less than in the 1946 volume.

The consideration of the best means of adapting the *Proceedings* to meet the need, mentioned in last year's Report, for rapid publication of important papers from leading British and Continental laboratories was continued during the year. An Advisory Board has been appointed to assist in the production of the *Proceedings*, which, beginning with Volume 60 (1948), will appear, without change of title, in two volumes per annum, each volume consisting of six parts issued monthly. In order to minimize the period between the communication and the publication of a paper, the monthly issues will be as large as the paper supply permits ; the period will probably be considerably shorter for papers of less than 3500 words than for longer papers.

REPORTS ON PROGRESS IN PHYSICS

Good progress has been made in the preparation of Volume 11 (1946-47), which should be published early in the summer of 1948. Throughout the year there was a satisfactory demand for copies of Volume 10 (1944-45) and the reprinted Volume 4 (1937), which are now the only volumes still obtainable at the office of the Society.

SPECIAL REPORTS

Two important conference reports, the active preparation of which was mentioned last year, have been completed and published. The Report of the Conference (April 1946) on *Meteorological Factors in Radio-wave Propagation* appeared in June 1947, and sales of copies were increasing satisfactorily by the end of the year. The Report of the International Conference at Cambridge (July 1946) was published in December in two volumes : Volume I, *Fundamental Particles* ; Volume II, *Low Temperatures* ; the number of early requests for copies is most encouraging.

The *Report on Colour Terminology*, on the preparation of which a Sub-Committee of the Colour Group has been engaged for several years, is now in the press and should be ready early in 1948.

It will be recalled that the Society has already published, as special parts of the *Proceedings*, reports of two summer conferences held in the H. H. Wills Physical Laboratory of the University of Bristol, namely those on "The Conduction of Electricity in Solids", in 1937, and "Internal Strains in Solids", in 1939. In continuation of the series, the report of a third conference, on "Strength of Solids", in July 1947, is now in active preparation as a Special Publication which should be ready early in the summer of 1948.

By arrangement with the Royal Society, the report of an International Conference on "The Emission Spectra of the Night Sky and Aurorae", organized by the Gassiot Committee in June 1947, is to be published by the Physical Society ; it also should appear in the summer of 1948. An earlier Report of the Gassiot Committee was included in *Reports on Progress in Physics*, Volume 9 (1942-43).

REPRESENTATION ON OTHER BODIES

The following appointments of representatives in 1947 are reported :—

Board of the Institute of Physics : Dr. H. H. Hopkins, Dr. A. C. G. Menzies.

Joint Publications Committee of the Physical Society and the Institute of Physics : Professor G. I. Finch, Professor E. N. da C. Andrade, Dr. H. H. Hopkins, Dr. C. H. Collie, Dr. W. S. Stiles, Dr. W. D. Wright, Mr. E. W. H. Selwyn, Mr. J. H. Awbery, Mr. A. T. Pickles.

British Committee for Radiological Units : Professor H. T. Flint, Sir John Cockcroft, Mr. W. G. Marley, Mr. D. W. Fry, Professor E. A. Owen.

B.S.I. Committee for X-ray Diffraction Powders : Mr. C. Wainwright.

Illuminating Engineering Society, Sound-film Projector Committee : Dr. R. S. Clay.

Royal Society Committee on Accommodation for Scientific Societies : Professor G. I. Finch.
National Committee for Theoretical and Applied Mechanics : Mr. C. N. H. Lock.
Royal Society Committee on Symbols : Mr. J. H. Awbery and Professor H. T. Flint.

OBITUARY

The Council records with great regret the deaths of the following: Lord Rayleigh, *Past President*; Professors F. Paschen and Max Planck, *Honorary Fellows*; Professor E. van Aubel, Mr. R. H. Dickinson, Professor P. N. Ghosh, Major M. van S. Godfery, Mr. W. Green, Dr. W. J. H. Moll, Dr. A. Müller, Mr. G. T. Purves, Mr. A. A. Smith, Mr. A. P. Trotter and Mr. P. C. Vincent, *Fellows*; and Mr. W. B. Richardson and Mr. J. H. Strong, *Student Members*.

The Society was represented at the funeral of Lord Rayleigh at Terling Church, Essex, by Dr. W. Jevons and Professor A. O. Rankine, and at the memorial service at St. Margaret's Church, Westminster, by Professor L. C. Martin and Dr. H. Shaw.

MEMBERSHIP

As the following tables show, the notable expansion of the Society in recent years has been not merely maintained but even accelerated. The total membership showed a bigger increase than in any previous year. The number of newly elected Student Members slightly exceeded the previous highest, in 1942. The number of newly elected Fellows has only once been exceeded, in 1946; but the transfers from Student Membership were fewer than in 1945 and 1946.

Roll of Membership		Hon. Fellows	Ex- officio Fellows	Fellows	Student Members	Total
Totals, 31 Dec. 1946		10	4	1332	271	1617
Changes during 1947	Newly elected	2		126	102	
	Transferred			33	33	
	Deceased			12	2	
	Resigned Lapsed			10	11	
Net increase		-2		137	44	179
Totals, 31 Dec. 1947		8	4	1469	315	1796

Year	1940	1941	1942	1943	1944	1945	1946	1947
Newly elected Student Members .	39	51	96	74	52	68	53	102
Transfers from Student Membership	6	13	20	22	23	52	46	33
Newly elected Fellows	20	28	57	42	56	56	139	126
Net increase in Fellowship . .	-25	6	57	36	55	80	165	137
Net increase in total Membership .	-14	37	140	73	63	75	160	179

GROUPS

COLOUR GROUP

The Group held its Sixth Annual General Meeting in the rooms of the Royal Photographic Society on 26 March 1947, when the Committee for 1947-48 was elected, Mr. J. G. Holmes being elected as Chairman and Dr. W. D. Wright re-elected as Honorary Secretary.

During the year there were five Science Meetings, brief particulars of which are given in

Report of Council

the *Proceedings**. One of them, at Imperial College on 1 October, was devoted to a review and discussion of an International Conference on Colour Vision which was held at Cambridge in July 1947. Several of the papers read at meetings of the Group have been published in the *Proceedings*.†

The Report of the Sub-Committee on "Colour Terminology" is now in the press, and will shortly be published as a companion to the earlier Report on "Defective Colour Vision in Industry".

OPTICAL GROUP

At the Fifth Annual General Meeting, which was held at Imperial College on 9 May 1947, Professor L. C. Martin was elected as Chairman of the Group, Mr. E. W. H. Selwyn was re-elected as Honorary Secretary and the Committee for 1947-48 was elected.

Six Science Meetings took place in 1947, and are briefly reported in the *Proceedings**. At one of them, at Imperial College on 14 February, the work of the Réunion d'Opticiens, held in Paris in October 1946, was reviewed and discussed. Several of the papers read at the meetings of the Group have been published in the *Proceedings*‡, and some in the *Photographic Journal*.

LOW-TEMPERATURE GROUP

In 1947 the Group held five Science Meetings, three of which were at the Science Museum and two took the form of visits for inspection and demonstration of the work of the Clarendon Laboratory, Oxford, and Messrs. J. and E. Hall Ltd., Dartford, Kent. Brief records of the meetings are given in the *Proceedings**.

The Second Annual General Meeting of the Group was held at the Science Museum on 18 December 1947, when Sir Charles Darwin and Professor F. E. Simon were elected as Chairman and Vice-Chairman respectively, Dr. G. G. Haselden was re-elected as Honorary Secretary, and the Committee for 1948 was elected.

ACOUSTICS GROUP

The Inaugural General Meeting of the Group took place at the Royal Institute of British Architects on 19 February 1947 for the consideration and adoption of a draft Constitution and for the election of Officers and Committee for 1947-48. Mr. H. L. Kirke and Dr. Alex Wood were elected as Chairman and Vice-Chairman, respectively, and Mr. W. A. Allen and Mr. A. T. Pickles as Joint Honorary Secretaries. An Inaugural Address on "The Contribution of Acoustical Science to Allied Studies" was delivered by Dr. Alex Wood.

Three Science Meetings, brief particulars of which are given in the *Proceedings**, were held during the year. The first of these was devoted to a two-day symposium on Reverberation, an account of which will appear in due course as a Special Publication of the Society.

MEMBERSHIP

The membership of the four Groups on 31 December 1947 was as follows :—

	Colour	Optical	Low-Temperature	Acoustics
Members of the Physical Society .	112	179	44	53
Members of participating bodies .	64	70	35	103
Members of subscribing firms .	18	23	—	—
Other members	16	14	9	49
	<u>210</u>	<u>286</u>	<u>88</u>	<u>205</u>

* *Proceedings*, 1947, 59, xi-xv ; 1948, 61, vi-ix.

† *Proceedings*, 1947, 59, 554, 560, 574, 592.

‡ *Proceedings*, 1947, 59, 44, 155, 940.

REPORT OF THE HONORARY TREASURER FOR THE YEAR ENDED 31 DECEMBER 1947

The total expenditure (£17,504) and the total income (£15,511) were higher than in 1946 by £3590 and £3081 respectively, and show a deficit on the year of £1992 7s. 3d.

The large increase (£2949) which again occurs in the cost of publications is due partly to the rising costs of paper and printing, but mainly to the Special Reports on Meteorological Factors in Radio Wave Propagation and an International Conference on Fundamental Particles and Low Temperatures, held in Cambridge in 1946, which were produced during the year. A further appreciable increase also occurs in the cost of *Science Abstracts*.

On the income side the steady advance in receipts from subscriptions continues, and this item should record a further large increase during 1948. Sales of *Proceedings* have risen by £1124, due mainly to an increased demand by libraries.

A publications grant of £1000 from the Royal Society is gratefully acknowledged, together with a further amount of £400 as a special grant towards the cost of publishing the Gassiot Conference Report.

At the auditors' recommendation the Society's stock of publications (£6700, kindly valued by Messrs. Wm. Dawson and Sons Ltd.) has been included and has accordingly been taken direct to the General Fund on one side of the balance sheet, and shown as an asset on the other side. Next year this figure will appear on the debit side of the Income and Expenditure Account with the corresponding figure for Stock at the close of the year, thereby giving a true result of the year's operations.

In the Balance Sheet "subscriptions in advance" record a considerable increase, due to the receipt in December of a very much larger number of subscriptions for the following year, than has been the case in past years. Investments decreased by £1400, due to the sale of Lancaster Corporation Stock (£400) and 3% Savings Bonds (£1000). In addition, the sum of £900 was withdrawn during the year from the Post Office Savings Bank.

(Signed) H. SHAW,
Honorary Treasurer.

12 February 1948.

SPECIAL FUNDS

W. F. STANLEY TRUST FUND

	£	s.	d.		£	s.	d.
Carried to Balance Sheet	259	0	0	£300 Southern Railway Preferred Ordinary Stock	199	0	0
				£442 Southern Railway Deferred Ordinary Stock	60	0	0
	£259	0	0		£259	0	0

DUDELL MEMORIAL TRUST FUND

CAPITAL

	£	s.	d.		£	s.	d.
Carried to Balance Sheet	374	0	0	£400 3½% War Loan Inscribed "B" Account	374	0	0

REVENUE

	£	s.	d.		£	s.	d.
Certificate	1	0	10	Balance on 31 December 1946	1	0	9
Balance carried to Balance Sheet	13	19	11	Interest on War Loan	14	0	0
	15	0	9		£15	0	9

SPECIAL FUNDS (*contd.*)

CHARLES CHREE MEDAL AND PRIZE FUND

CAPITAL

	£	s.	d.		£	s.	d.
Balance carried to Balance Sheet . . .	1866	10	0	£784 4% Funding Stock	839	0	0
				£1500 2½% Consols	1027	10	0
	1866	10	0		1866	10	0

REVENUE

	£	s.	d.		£	s.	d.
Prize and Certificate	132	5	10	Balance on 31 December 1946	81	3	4
Balance carried to Balance Sheet . . .	17	14	8	Interest on Investments	68	17	2
	£150	0	6		£150	0	6

CHARLES VERNON BOYS PRIZE FUND

CAPITAL

	£	s.	d.		£	s.	d.
Balance carried to Balance Sheet . . .	900	0	0	£1132 16s. 10d. 2½% Consols "B" Account	900	0	0

REVENUE

	£	s.	d.		£	s.	d.
Balance on 31 December 1946	16	10	9	Interest on Investment	28	6	4
Prize	26	5	0	Balance carried to Balance Sheet	15	10	3
Certificate	1	0	10				
	£43	16	7		£43	16	7

HOLWECK PRIZE FUND

CAPITAL

	£	s.	d.		£	s.	d.
Balance carried to Balance Sheet . . .	675	0	0	£675 3% Defence Bonds	675	0	0

REVENUE

	£	s.	d.		£	s.	d.
Prize and Certificate	107	0	0	Balance on 31 December 1946	21	12	2
Expenses	56	6	9	Donation	3	0	0
Balance carried to Balance Sheet . . .	8	13	8	Interest on Investment	22	8	3
				Sale of Investment	100	0	0
				Grant from General Fund	25	0	0
	£172	0	5		£172	0	5

ADDENBROOKE BEQUEST

CAPITAL

	£	s.	d.		£	s.	d.
Balance carried to Balance Sheet . . .	337	0	0	£384 6s. 7d. 2½% Consols "D" Account	337	0	0

REVENUE

	£	s.	d.		£	s.	d.
Balance carried to Balance Sheet . . .	16	13	0	Balance on 31 December 1946	6	0	6
				Interest on Investment	9	12	0
				Refund	1	0	6
	£16	13	0		£16	13	0

LIFE COMPOSITION FUND ON 31 DECEMBER 1947

	£	s.	d.
17 Fellows paid £10	170	0	0
1 Fellow paid £15	15	0	0
18 Fellows paid £21	378	0	0
1 Fellow paid £25 4s. 0d.	25	4	0
1 Fellow paid £26 5s. 0d.	26	5	0
57 Fellows paid £31 10s. 0d.	1795	10	0
1 Fellow paid £33 12s. 0d.	33	12	0
	<u>£2443</u>	<u>11</u>	<u>0</u>

INCOME AND EXPENDITURE ACCOUNT FOR THE YEAR ENDED 31 DECEMBER 1947

[illegible]

BALANCE SHEET AS ON 31 DECEMBER 1947

LIABILITIES		£	s.	d.
<i>Sundry Creditors</i>	.	1773	12	8
<i>Bank Overdraft</i>	.	111	5	10
<i>Life Compositions:</i>				
As on 31 December 1946	.	2357	10	0
Add Payments during year	.	127	1	0
		2484	11	0
Less Transfer to Income and Expenditure Account	.	41	0	0
		2443	11	0
<i>Subscriptions received in advance:</i>				
Members	.	694	8	8
Publications	.	731	8	6
		1425	17	2
<i>SPECIAL FUNDS:</i>				
<i>W. F. Stanley Trust Fund</i>	.	259	0	0
<i>Duddell Memorial Trust Fund</i>	.	387	19	11
<i>Charles Chree Medal and Prize Fund</i>	.	1884	4	8
<i>Charles Vernon Boys Prize Fund</i>	.	884	9	9
<i>Holweck Prize Fund</i>	.	683	13	8
<i>Addenbrooke Bequest</i>	.	353	13	0
		4453	1	0
<i>GENERAL FUND:</i>				
As on 31 December 1946	.	6273	11	2
Add Value of Stock of Publications on 31 December 1947	.	6700	0	0
Profit on Sale of Investments	.	31	17	6
		13,005	8	8
Less Balance of Income and Expenditure Account	.	1992	7	3
		11013	1	5

We have audited the above Balance Sheet and have obtained all the information and explanations we have required. We have verified the bank balances and the investments. In our opinion such Balance Sheet is properly drawn up so as to exhibit a true and correct view of the state of the Society's affairs according to the best of our information and the explanations given to us and as shown by the books of the Society.

KNOX, CROPPER & Co.,
Chartered Accountants.

SPENCER HOUSE, SOUTH PLACE, E.C.2.
12th February 1948.

£21,220 9 1

ASSETS		£	s.	d.	£	s.	d.
<i>Investments at market value on 31 December 1939 or cost:—</i>							
<i>W. F. STANLEY TRUST FUND:</i>							
£300 Southern Railway Preferred Ordinary Stock.		199	0	0			
£442 Southern Railway Deferred Ordinary Stock.		60	0	0			
		259	0	0			
<i>DUDELL MEMORIAL FUND:</i>							
£400 3½% War Stock	.				374	0	0
<i>CHARLES CHREE MEDAL AND PRIZE FUND:</i>							
£784 4% Funding Stock	.	839	0	0			
£1500 2½% Consols	.	1027	10	0			
		1866	10	0			
<i>CHARLES VERNON BOYS PRIZE FUND:</i>							
£1132 16s. 10d. 2½% Consols	.				900	0	0
<i>HOLWECK PRIZE FUND:</i>							
£675 3% Defence Bonds	.				675	0	0
<i>ADDENBROOKE BEQUEST:</i>							
£384 6s. 7d. 2½% Consols	.				337	0	0
<i>GENERAL FUND:</i>							
£2100 2½% Consols	.	1437	10	0			
£1750 3½% War Stock	.	1636	0	0			
£500 4% Funding Stock	.	535	0	0			
£399 L.M.S. 4% Debentures	.	355	0	0			
£1000 L.M.S. 4% Preference Stock.	.	595	0	0			
£500 L.N.E.R. 4% Debentures	.	412	0	0			
£150 S.R. 5% Debentures	.	171	0	0			
£2200 3% Defence Bonds	.	2200	0	0			
		7341	10	0			
		11,753	0	0*			
* Market value on 31 December 1947: £13,106.							
<i>Dividends due on Investments</i>					95	7	3
<i>Inland Revenue—Income Tax recoverable for 1946</i>					43	12	8
<i>Subscriptions due</i>					77	1	8
<i>Sundry Debtors</i>					1050	6	3
<i>Stock of Paper and Binding Material</i>					675	7	6
<i>Publications</i>					6700	0	0
<i>32nd Exhibition (1948)—Expenditure to date</i>					7375	7	6
<i>Cash in Post Office Savings Bank Account</i>					663	13	3
<i>Cash in hand</i>					93	9	5
					68	11	1
					162	0	6
					£21,220	9	1

THE PROCEEDINGS OF THE PHYSICAL SOCIETY

VOL. 61, PART 1

1 July 1948

No. 343

The Atomic Heats of Palladium, Sodium and Mercury at Low Temperatures

By G. L. PICKARD* AND F. E. SIMON

Clarendon Laboratory, Oxford

* Now at University of British Columbia, Vancouver

MS. received 5 March 1948

ABSTRACT. The specific heats of palladium and sodium were measured between helium and hydrogen temperatures, the specific heats of mercury between helium and liquid air temperatures. Palladium shows a very high electronic term of the specific heat. Sodium exhibits anomalous behaviour in various respects. In the case of mercury a search was made for a transition at low temperatures predicted by Bridgman, but no indication was found.

The thermodynamic functions of the three metals between absolute zero and room temperature are given in tables.

§1. INTRODUCTION

THE object of this series of measurements was to extend the known data on the atomic heats of these metals for the following individual reasons. For palladium it was desired to determine the value of the electronic specific heat below 20°K. In the case of sodium an anomalous behaviour had been suspected (Simon 1926) and experiments by Kurti and Simon using the magnetic cooling method had also suggested that the atomic heat in this region was abnormally high. It was therefore considered desirable to measure the specific heat of sodium in the helium region. For mercury the possibility of a polymorphic transition manifesting itself as a discontinuity in the atomic heat/temperature curve was to be investigated.

The experimental method in each case was to use a vacuum calorimeter with electrical heating and temperature measurement. The calorimeter was enclosed within a vacuum case in a small helium "expansion" liquefier, the sample being cooled by the condensation into the calorimeter of liquid helium which was then pumped away and the calorimetric measurements made over the required temperature range.

Since the method used was of a conventional character only an outline of the experimental method is given. The cylindrical calorimeters used were of copper of 0.2 mm. thickness, tinned to avoid leaks and supported by a 2 mm. diameter thin-walled german silver tube through which the helium was condensed and pumped. The calorimeter for the palladium measurements was 1.7 cm. in diameter and 2.1 cm. long. The electrical thermometer was of silk insulated 47 s.w.g. Eureka wire of 535 Ω resistance while the heater coil of 550 Ω resistance was of the same wire. These coils were wound round the outside of the calorimeter and fixed in position by painting with bakelite varnish which was hardened

by baking for four hours. The resistance of the thermometer coil was designed to suit the Wheatstone bridge and galvanometer available and enabled the temperature to be measured to $\pm 0.02^\circ$ below 20°K . The 30 s.w.g. Eureka leads passed through a heat trap connected to the helium expansion chamber to reduce heat conduction to the calorimeter.

The second calorimeter was 3.0 cm. in diameter and 3.8 cm. long to accommodate the larger specimens of mercury and sodium. The heater coil of 40 s.w.g. Eureka was of $180\ \Omega$ resistance and the 47 s.w.g. Eureka thermometer coil was of $485\ \Omega$ resistance. As measurements with the mercury were also required at temperatures well above 20°K ., where the Eureka thermometer becomes insensitive, it was supplemented by a platinum resistance thermometer of $40\ \mu$ bare wire wound on a thin paper base and attached to the calorimeter by bakelite varnish. This coil had a resistance of $322\ \Omega$ at 273°K ., falling to $4.5\ \Omega$ at 25°K . Arrangements were made to measure the lead resistance to the platinum coil. The temperature corresponding to the platinum resistance was determined from tables published by Clusius and Vaughen (1929). The wire used in the present experiments followed closely the values for the sample Pt-23'-1915 quoted in those tables. The Eureka thermometers were calibrated in the hydrogen and helium regions by means of the vapour pressure/temperature data for these liquids, and in the intermediate region an interpolation was made on the basis of the information obtained in a previous experiment (Pickard and Simon 1948).

The samples were cooled to about 10°K . by condensing hydrogen into the calorimeter and then pumping it away. They were further cooled by condensing helium into the calorimeter and then pumping it. When all this helium had been removed, a small quantity of helium gas was re-introduced to permit thermal exchange to take place between the calorimeter and the sample. When the temperature drift of the calorimeter had become steady the calorimetric measurements were begun. The temperature was measured at 20-sec. intervals throughout this stage of the experiments, heat being supplied for periods of one or two minutes with intervals to allow the temperature drift to be measured. The experiments were carried out in such a way that the temperature difference between the calorimeter and its enclosure was always small and the thermal isolation of the calorimeter was good. It was therefore unnecessary to use any special method to correct for imperfect heat insulation. The rise in temperature due to the heating was determined by plotting a temperature/time curve before and after the heating and extrapolating to the middle of the heating period, thus correcting for the temperature drift.

During the heating the temperature of the calorimeter rose above that of its contents, but thermal equilibrium was rapidly attained after switching off the heater current. It was found that the use of different heater currents introduced no systematic error into the results. The "overheating" was of the order of $\frac{1}{2}^\circ$ with the mercury sample; with the palladium it was less than this, and with the sodium it was so small as to be hardly noticeable.

Correction was made for the heat capacity of the calorimeter using the values for the specific heat of copper and tin measured at Leiden (Kok and Keesom 1936, Keesom and Ende 1932). The heat capacity of the resistance coils was negligible and that of the helium exchange gas could be neglected at all except the lowest temperatures.

The metals will now be dealt with individually, any special details in technique being described with reference to the particular metal concerned.

§ 2. PALLADIUM

Keesom and Clark (1935) have shown that the atomic heat of nickel at low temperatures contains a linear term $0.001744 T$ which is many times bigger than the contribution from free electrons to be expected from the Sommerfeld theory. Mott (1936) has suggested that this term is due to the unfilled quantum states or positive holes in the d -shell of nickel, and that a similar term should occur for the other transition metals, whether ferromagnetic or paramagnetic.

In order to test this theory and to make more quantitative data available on the subject, the specific heat of palladium was determined, palladium being chosen on account of its similarity to nickel and its high paramagnetism which, according to Mott, should imply a large electronic specific heat. Though the electronic component in these transition metals is larger than usual, it is nevertheless small

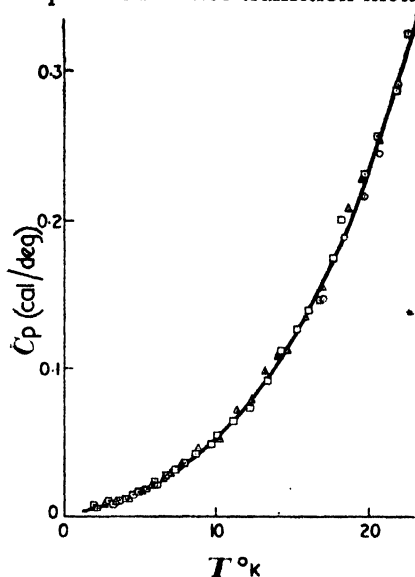


Figure 1. Atomic heat of palladium.

Curve represents equation

$$C = 0.0000224 T^3 + 0.0031 T.$$

○ △ □ Pickard and Simon.

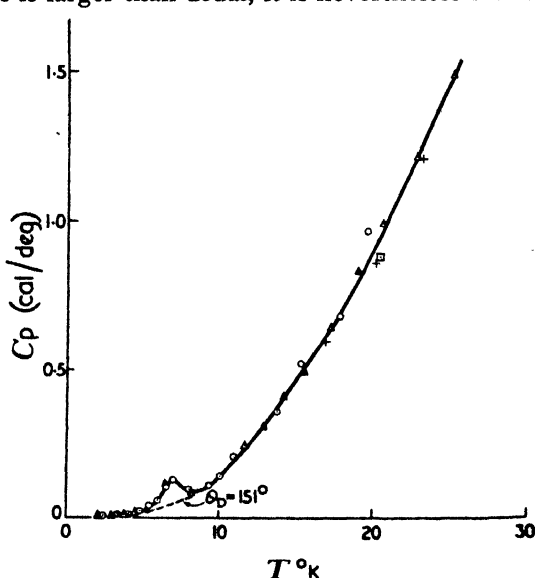


Figure 2. Atomic heat of sodium.

○ △ Pickard and Simon ;

+ Simon and Zeidler.

in absolute magnitude. But since it varies in direct proportion to the temperature, whereas the specific heat due to the lattice vibrations varies near the absolute zero with the cube of the temperature, the two become comparable in magnitude at sufficiently low temperatures as pointed out by Simon (1928). Experiments were therefore carried out in the liquid hydrogen/helium region.

The sample of palladium was kindly put at our disposal by Professor N. F. Mott and was a solid cylinder of the metal, 1.4 cm. in diameter and 1.4 cm. long, weighing 25.21 gm. The palladium sample was strongly heated *in vacuo* to determine if it contained any absorbed hydrogen. The amount obtained was scarcely detectable and certainly too small to affect the results at all. Its atomic heat was measured in the small calorimeter in the range 2–22° K.

The values obtained for the atomic heat are plotted in figure 1. A mean curve was drawn and was analysed by trying a series of values of the Debye

characteristic temperature (θ_D) to calculate the lattice atomic heat (C_L) and finding which gave the most nearly linear remainder for the electronic specific heat (C_E). The best fit was found with $\theta_D = 275^\circ$, and table 1 gives values of

Table 1. Atomic heat of palladium (in cal/deg. gm. atom)

Temp. ($^\circ$ K.)	C_{obs}	$C_L(\rho=275^\circ)$	C_E/T
2	0.0066	0.0002	0.0032
4	0.0132	0.0014	0.00295
6	0.023	0.0046	0.0031
8	0.036	0.0114	0.0031
10	0.054	0.0224	0.00315
12	0.076	0.039	0.0031
14	0.106	0.062	0.00315
16	0.142	0.0915	0.00315
18	0.186	0.130	0.0031
20	0.240	0.178	0.0031
22	0.306	0.237	0.00315

the observed total atomic heat (C_{obs}) from the experimental curve, with values of C_L for $\theta = 275^\circ$ and of C_E/T . The values actually measured are of the atomic heat at constant pressure, but since the correction $C_p - C_v$ is very small at these low temperatures, the observed values have been taken as C_v .

The observed atomic heat can therefore be represented as the sum of a T^3 term (lattice atomic heat) and a T term (electronic specific heat) in the range 2° – 22° K. as

$$C = 0.0000224 T^3 + 0.0031 T \text{ cal/deg. gm. atom.}$$

In figure 1 a curve of this form is plotted for comparison with the experimental points.

The value of $\theta_D = 275^\circ$ from the present results may be compared with the value of $\theta_D = 270^\circ$ given by Meissner and Voigt (1930) from electrical measurements and with that of 260° from the Lindemann melting point formula.

In the expression for the atomic heat at 2.5° K. the linear term is about twenty times as great as the T^3 term while the two become of equal magnitude at about 12° K. The value for the electronic specific heat is, as predicted by Mott, much higher than that of nickel ($C_E = 0.001744 T$) while work by Kok and Keesom (1936) has shown that another transition metal, platinum, also exhibits a large electronic specific heat which, however, is not as great ($C_E = 0.001607 T$) as that for palladium.

A short note on these experiments has been published in *Nature* (Pickard 1936).

A short time ago Clusius and Schachinger (1947) measured the specific heat of palladium between 14° K. and room temperature and they found very good agreement with our experiments in the overlapping region. In addition, these authors have now all the data available to analyse the whole specific heat curve (including measurements of other authors up to 1000° K.), and in combining their values with those of the paramagnetic susceptibility they are now in a position to estimate the electronic specific heat over the whole temperature range between absolute zero and 1000° .

§ 3. SODIUM

The sodium used was from a sample of pure metal, hydrogen free, which had been prepared and used for experiments with neutrons in this laboratory. 18.90 gm. of sodium in the form of a cylinder which just fitted within the large calorimeter was used and measurements were made in the range 2–25° K.

The experimental points are shown graphically in figure 2 while some values for the atomic heat are given in table 2.

Table 2. Atomic heat of sodium

Temp. (° K.)	C_v	θ_D (°)	Temp. (° K.)	C_v	θ_D (°)
2	0.0055	88	10	0.140	149
3	0.0080	116	12	0.250	147
4	0.0135	130	14	0.385	148
5	0.027	129	16	0.54	150
6	0.059	120	18	0.71	152
7	0.124	109	20	0.90	154
8	0.093	137	25	1.45	156
9	0.100	150			

The present values are slightly higher than those of Simon and Zeidler (1926), who measured the specific heats between 17 and 60° K. Their values, which are shown in figure 2 together with the new ones, correspond to a θ of 159°. It can be seen from table 2 that the characteristic temperature becomes progressively lower with falling temperature.

Doubt had been expressed by Simon (1926) whether the value of θ of about 160° really represented the characteristic temperature due to lattice vibrations. He was led to this suggestion by the fact that the characteristic temperatures derived from Lindemann's melting point formula and from electrical conductivities were considerably higher. In particular, the latter values could be interpreted on the basis of a θ of 208°. Simon had suggested that, as in the case of grey tin and other substances discussed in his paper, this specific heat curve should be interpreted by a Debye term ($\theta = 202^\circ$) plus a "Schottky" term ($\theta = 95^\circ$) due to an internal transition of as yet unknown nature. Later, when Blackman (1934, 1935) had shown that the Debye function was less well founded than had previously been assumed, the opinion had been expressed that there was no need to assume the existence of such internal transitions, and the incompatibility of the specific heat values with the Debye curve was attributed to the defects of the latter theory.

While it is certainly possible that some relatively small anomalies in specific heat curves are due to this, we maintain the opinion that in the majority of cases discussed in the paper mentioned above, we are faced with internal transitions. A systematic investigation of the whole question has now been started at the Clarendon Laboratory, including measurements of thermal expansion, which will provide an independent value of θ . In the case of sodium, recent very careful experiments by MacDonald and Mendelssohn (1948) had shown that the electrical conductivity down to 2° K. can be represented very accurately with a θ of 202°. The difference between this value and the one derived from the specific heats without the assumption of an anomaly is much too big to be explained by a deficiency of the Debye theory.

The most striking feature in the specific heat curve is the anomaly at about 7° (see figure 3). An anomaly of a similar type had been observed by Cristescu

and Simon (1934) in beryllium at 11° . In both cases the total energy due to this anomaly is not of the order RT , but in the case of sodium $RT/70$ and in the case of beryllium $RT/700$. These are of the same order as the anomalies in the specific heat observed in superconductors, and investigation of the electrical properties (see also Kurti and Simon 1935) may shed more light on the nature of this anomaly.

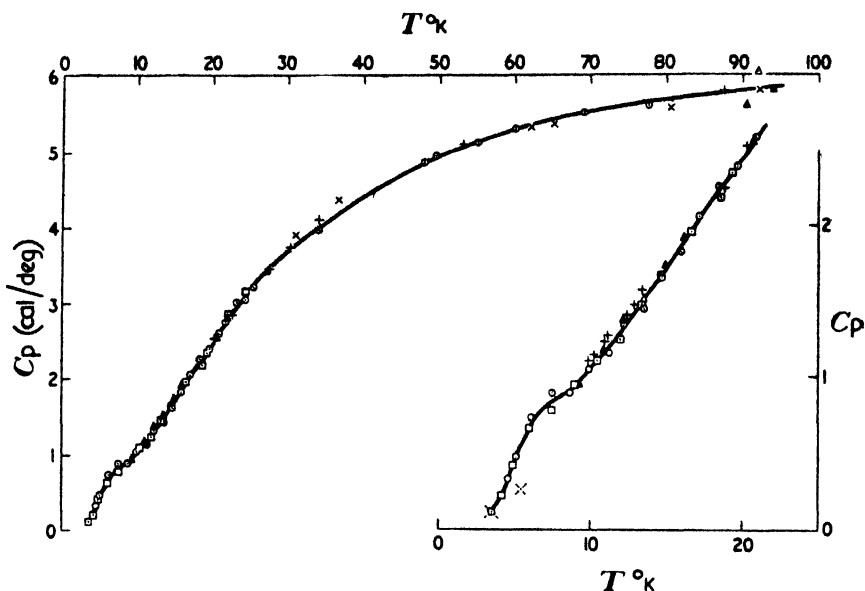


Figure 3. Atomic heat of mercury.

○ △ □ Pickard and Simon ; + Simon ; × Pollitzer ; × Onnes and Holst.

§4. MERCURY

The object of the experiments was to investigate the possibility that a polymorphic transition, as predicted by Bridgman (1935), might occur in the region below liquid oxygen temperatures. It would manifest itself in the form of a discontinuity in the total heat/temperature curve. It was at the same time desired to extend the previous measurements (Pollitzer 1911, 1913, Simon 1923 a, 1923 b) down to the helium region.

Bridgman's measurements were concerned with the polymorphic transitions of elements at very high pressures. An extrapolation of the data obtained by him for the temperature variation of the transition pressure between the polymorphic forms of mercury indicated that this transition might occur at normal pressure at a low temperature. From an inspection of Simon's measurements (1923 a, 1923 b) he claimed that a distinct break in the tangent to the atomic heat/temperature curve was evident between 53° – $68^\circ K$. and that this might be evidence of the occurrence of the transition.

The rather complicated crystal structure of mercury makes it improbable that its specific heat could be represented by a single characteristic temperature; Simon (1923 b) had found good agreement with the experimental data, using a two-term expression consisting of a Debye-function and an Einstein-function, the latter having a very low characteristic temperature, and, therefore, being responsible for the hump in the solid hydrogen region. A change in the curvature

was therefore to be expected. In view of Bridgman's prediction of a transition,¹ however, it seemed desirable to check the specific heat again in the region between hydrogen and air temperatures.

The method adopted in the search for any thermal discontinuity was to heat the calorimeter slowly and continuously by passing a constant current through the heater coil, and to measure and record the temperature at the usual 20-sec. intervals. A graph of temperature against time was then plotted and examined for evidence of any sudden change which would have been indicated by a discontinuity in the curve.

The sample of very pure mercury (75.80 gm.) was kindly lent to us by Dr. H. London and was contained in a vessel of thin copper, varnished to prevent amalgamation of the mercury with the copper. This vessel was placed in the large calorimeter. The measurements of the atomic heat were extended over the range from 3.5° to 95° K. The slow continuous heatings extended over the whole region from 25°–80° K.

The experimental points for the atomic heat are given in figure 2 and a mean curve is drawn through them. Some values from this mean curve are given in table 3.

Table 3. Atomic heat of mercury (C_v)

Temp. (° K.)	C_v	Temp. (° K.)	C_v
3.5	0.12	25	3.18
4	0.20 ₆	30	3.68
6	0.68	40	4.41
8	0.87 ₅	50	4.95
10	1.06	60	5.30
12	1.31	70	5.52
15	1.72	80	5.67
20	2.46	90	5.80

The present values for the atomic heat between 10° and 14° K. and above 25° K. are slightly lower than the previous ones of Simon (1923 a, 1923 b) which are shown in figure 3 together with the results of Pollitzer (1911, 1913) and two early results of Kamerlingh Onnes and Holst (1914). The present measurements confirm that the atomic heat is anomalously high at low temperatures. The results employing continuous heating were carefully examined but *no* evidence of any thermal discontinuity was observed within the limits of the experimental accuracy, about 0.02 cal/gm. atom in the region investigated. It must be mentioned that although the temperature was only *recorded* every 20 sec. it was actually measured almost continuously, and so the possibility of even a very small discontinuity passing unnoticed is very remote.

It should be added that mercury becomes superconducting at 4.2° and that there is a small discontinuity in the specific heat at this point (Daunt and Mendelssohn 1937).

§ 5. THERMODYNAMIC FUNCTIONS

Our measurements at the lowest temperatures together with those of other observers at higher temperatures now permit calculation of the thermodynamic data of the three metals. Tables 4 to 6 present in the usual manner the specific heat (C_p) entropy (S), enthalpy (I) and thermodynamic potential at constant pressure (F_p) at intervals of 10°. We are very much obliged to Mr. G. K. White for the trouble he has taken in calculating these data.

Table 4. Palladium (1 gm. atom = 106.7 gm.)

$T(^{\circ}\text{K.})$	C_p	$S-S_0$	$I-I_0$	$-(F_p-F_{p_0})$	$T(^{\circ}\text{K.})$	C_p	$S-S_0$	$I-I_0$	$-(F_p-F_{p_0})$
0	0.0	0.0	0.0	0.0	160	5.43	5.39	491	372
10	0.054	0.038	0.110	0.27	170	5.53	5.72	545	428
20	0.24	0.121	1.40	1.02	180	5.63	6.04	601	486
30	0.66	0.30	5.71	3.27	190	5.72	6.35	658	548
40	1.30	0.57	15.3	7.54	200	5.79	6.64	715	613
50	1.93	0.93	31.2	15.15	210	5.86	6.93	774	681
60	2.56	1.34	53.6	26.5	220	5.92	7.20	833	752
70	3.10	1.77	81.9	42.1	230	5.97	7.47	892	825
80	3.56	2.22	115	62.1	240	6.02	7.72	952	901
90	3.94	2.66	153	86.1	250	6.06	7.97	1012	980
100	4.26	3.09	194	115	260	6.10	8.21	1073	1061
110	4.54	3.51	238	148	270	6.14	8.44	1134	1144
120	4.79	3.92	285	185	280	6.18	8.66	1196	1230
130	5.00	4.31	334	226	290	6.22	8.88	1258	1317
140	5.16	4.68	385	271	300	6.27	9.09	1320	1407
150	5.30	5.05	437	320					

Table 5. Sodium (1 gm. atom = 23.0 gm.)

$T(^{\circ}\text{K.})$	C_p	$S-S_0$	$I-I_0$	$-(F_p-F_{p_0})$	$T(^{\circ}\text{K.})$	C_p	$S-S_0$	$I-I_0$	$-(F_p-F_{p_0})$
0	0.0	0.0	0.0	0.0	160	6.00	8.40	665	679
10	0.135	0.067	0.452	0.118	170	6.07	8.76	725	764
20	0.905	0.38	5.23	2.31	180	6.13	9.11	786	854
30	2.00	0.96	19.9	8.81	190	6.18	9.44	848	947
40	2.99	1.67	45.0	21.9	200	6.24	9.76	910	1043
50	3.82	2.44	79.4	44.6	210	6.30	10.07	972	1142
60	4.36	3.19	120.6	70.7	220	6.36	10.31	1036	1233
70	4.75	3.89	166.6	105.6	230	6.41	10.60	1100	1338
80	5.02	4.54	215	148	240	6.47	10.87	1164	1445
90	5.24	5.15	267	196	250	6.53	11.14	1229	1556
100	5.42	5.71	320	251	260	6.59	11.40	1294	1668
110	5.56	6.23	375	310	270	6.65	11.65	1360	1784
120	5.66	6.72	431	375	280	6.71	11.89	1437	1892
130	5.75	7.17	488	444	290	6.77	12.13	1505	2012
140	5.84	7.60	546	518	300	6.83	12.36	1573	2134
150	5.92	8.01	605	596					

Table 6. Mercury (1 gm. atom = 200.6 gm.)

$T(^{\circ}\text{K.})$	C_p	$S-S_0$	$I-I_0$	$-(F_p-F_{p_0})$	$T(^{\circ}\text{K.})$	C_p	$S-S_0$	$I-I_0$	$-(F_p-F_{p_0})$
0	0.0	0.0	0.0	0.0	170	6.31	12.36	847	1254
10	1.06	0.70	4.62	2.26	180	6.36	12.72	911	1379
20	2.49	1.85	22.2	14.7	190	6.41	13.06	974	1507
30	3.68	3.11	53.6	39.5	200	6.47	13.39	1039	1640
40	4.41	4.27	94.3	76.5	210	6.53	13.71	1104	1775
50	4.95	5.32	141	125	220	6.62	14.01	1169	1914
60	5.29	6.25	193	182	230	6.74	14.31	1236	2055
70	5.51	7.08	247	249	234.4*	6.78	14.43	1265	2116
80	5.67	7.83	303	324	234.4†	7.00	16.84	1820	2116
90	5.79	8.51	360	405	240	6.95	17.01	1854	2221
100	5.89	9.12	419	494	250	6.90	17.28	1924	2418
110	5.98	9.69	478	588	260	6.85	17.56	1991	2570
120	6.05	10.21	538	687	270	6.76	17.81	2058	2746
130	6.10	10.7	599	792	280	6.70	18.06	2115	2939
140	6.16	11.15	660	901	290	6.65	18.29	2192	3107
150	6.21	11.58	722	1014	300	6.65	18.52	2258	3292
160	6.26	11.98	784	1132					

* Solid.

† Liquid.

REFERENCES

- BLACKMAN, M., 1934, *Proc. Roy. Soc. A*, **148**, 365, 384 ; 1935, *Ibid.*, **149**, 117, 126.
BRIDGMAN, P. W., 1935, *Phys. Rev.*, **48**, 898.
CLUSIUS, K., and SCHACHINGER, L., 1947, *Z. Naturforschung*, **2 a**, 90.
CLUSIUS, K., and VAUGHEN, J. V., 1929, *Z. ges. Kalte. Ind.*, **36**, 215.
CRISTESCU, S., and SIMON, F., 1934, *Z. phys. Chem. B*, **25**, 273.
DAUNT, J. G., and MENDELSSOHN, K., 1937, *Proc. Roy. Soc. A*, **160**, 127.
KEESOM, W. H., and CLARK, C. W., 1935, *Physica*, **2**, 513.
KEESOM, W. H., and VAN DEN ENDE, J. N., 1932, *Leiden: Comm.*, 219 b.
KOK, J. A., and KEESOM, W. H., 1936, *Leiden Comm.*, 245 a ; 1936, *Physica*, **3**, 1035.
KURTI, N., and SIMON, F., 1935, *Proc. Roy. Soc. A*, **151**, 610.
MACDONALD, D. K. C., and MENDELSSOHN, K., 1948, *Nature, Lond.*, in press.
MEISSNER, W., and VOIGT, B., 1930, *Ann. Phys., Lpz.*, **7**, 893.
MOTT, N. F., 1936, *Proc. Roy. Soc. A*, **152**, 42.
ONNES, H. KAMERLINGH, and HOLST, G., 1914, *Leiden Comm.*, 142 c.
PICKARD, G. L., 1936, *Nature, Lond.*, **138**, 123.
PICKARD, G. L., and SIMON, F. E., 1948, *Proc. Phys. Soc.*, **60**, 405.
POLLITZER, F., 1911, *Z. Elektrochem.*, **17**, 5 ; 1913, *Ibid.*, **19**, 513.
SIMON, F., 1923 a, *Ann. Phys., Lpz.*, **68**, 241 ; 1923 b, *Z. phys. Chem. A*, **107**, 279 ;
1926, *Ber. Pr. Akad. Wiss.*, **33**, 477 ; 1928, *Z. Elektrochem.*, **34**, 530.
SIMON, F., and ZEIDLER, W., 1926, *Z. phys. Chem. A*, **123**, 383.

The Theory of the Propagation of Second Sound in Helium II

By R. B. DINGLE,

H. H. Wills Physical Laboratory, Royal Fort, Bristol

Communicated by N. F. Mott ; MS. received 20 January 1948

ABSTRACT. In this paper the conduction of heat in liquid helium II will be discussed, and the velocity of second sound, the attenuation, and reflection coefficients under various conditions will be obtained. The shape of the velocity-temperature curve will be discussed on the basis of Landau's theory of helium II, and Tisza's theory will be criticized.

§1. INTRODUCTION

THE experiments of Peshkov (1946) in Russia and Lane and co-workers (1947) in America have demonstrated beyond doubt the existence of a new type of sound wave in liquid helium II, one in which temperature fluctuations are the important phenomena and not pressure fluctuations as in the case of ordinary sound. The essential assumptions of the theories developed by Tisza (1938, 1940 and 1947) and Landau (1941, 1944) which account for second sound are as follows: (i) Helium II may, at least from a phenomenological standpoint, be divided into two parts. One of these parts, the so-called superfluid, is probably to be explained as some kind of condensed phase, perhaps the condensed phase to be expected below a certain temperature in a substance obeying Bose-Einstein statistics (London 1938, 1939). (ii) The entropy of one part, the so-called normal fluid, is greater than that of the superfluid. Kapitza (1941) has shown, by experiments in which the two fluids are separated by the use of a very narrow gap, that the entropy of the superfluid is very small: it may be zero. (iii) To these assumptions may be added a third, which is important in the theory of the flow of helium II. The normal fluid obeys the hydrodynamical equations appropriate to any other normal

fluid; thus its equations contain a coefficient of viscosity; but the superfluid obeys hydrodynamical equations appropriate to a fluid in which the curl of the velocity is everywhere zero, so that the motion of the superfluid is always irrotational. In particular, the tangential velocity of the normal fluid must vanish at a surface, whilst the tangential velocity of the superfluid need not vanish there.

In this paper we shall consider several problems in the theory of the propagation of heat and of second sound in helium II using the above assumptions; the problems are chosen both because of their importance in experimental investigations and also as illustrations of the use of the equations and boundary conditions which will be established in this paper.

§ 2. THE SHAPE AND SPEED OF ORDINARY "CONDUCTED" PULSES

Throughout this section, the word "conducted" will mean a flow of heat not involving fluid movements.

When Tisza (1938, 1940) predicted the existence of "temperature waves", he pointed out that these waves could easily be distinguished from ordinary "conduction waves" by the fact that the latter would be greatly attenuated, and would also have a velocity which would be proportional to the square root of the frequency, unlike the temperature waves, which should have a velocity independent of the frequency.

For suppose that T is the temperature excess over the initial uniform value, t the time, K the ordinary thermal conductivity (due to heat flow without fluid movements), and C the specific heat at constant volume. Then elementary arguments give the equation of heat conduction as $dT/dt = (K/\rho C)\nabla^2 T$, where ρ is the density. Suppose for simplicity that we are dealing with plane propagation along the z axis. Then this equation becomes $dT/dt = Dd^2T/dz^2$ where $D = K/\rho C$. If the source consists of a plane at $z=0$ undergoing sinusoidal temperature fluctuations of frequency $\omega/2\pi$, then $dT/dt = i\omega T$; $d^2T/dz^2 = i\omega T/D$; and $T \propto e^{i\omega t} \exp\{-z(1+i)\sqrt{(\omega/2D)}\}$.

Hence the distance in which the amplitude of temperature falls to $1/e$ of the temperature of the source is just one wavelength, and the apparent velocity is $\sqrt{(2D\omega)}$.

If on the other hand the "conduction wave" is emitted from the plane $z=0$ at the time $t=0$ as a sharp pulse, the appropriate solution of the equation is

$$T \propto \exp(-z^2/4tD)/\sqrt{t},$$

where $D = K/\rho C$. If the conduction effect is the only one of importance, or if all the temperature waves of the Tisza-Landau type are reflected back without loss, then the proportionality constant may be found by integration of T (the excess temperature) over all space. The result must clearly be independent of time except during the initial sharp emission of the pulse. Thus if Q is the quantity of heat in the pulse,

$$Q = C \int_{-\infty}^{\infty} T dz = \frac{AC}{\sqrt{t}} \int_{-\infty}^{\infty} \exp(-z^2/4tD) dz = AC\sqrt{(4\pi D)},$$

where A is a proportionality constant. Thus finally

$$T = \frac{Q}{C\sqrt{(4\pi tD)}} \exp(-z^2/4tD).$$

The maximum of the pulse, for a given value of z , is given by $(\partial T/\partial t)_z=0$; i.e. $t_{\max}=z^2/2D$ and $T_{\max}=0.242Q/Cz$. Thus the apparent velocity of the pulse depends on the distance z ; this is simply because the conduction effect is dispersive. The shape of this "conducted pulse" may be deduced from the above equations.

In problems of this type it is only necessary to obtain solutions of the physical problem involved for the simple case of pure sine waves; all other cases may be deduced mathematically from this solution by means of Fourier analysis.

§ 3. THE VELOCITY OF SECOND SOUND

A satisfactory deduction of the velocity of second sound was first given by Landau (1941). Recently a simplified approximate treatment has been proposed by Gogate and Pathak (1947). They obtain the same formula for the velocity, but some of their arguments appear unsatisfactory to the present author, especially as in the deduction of one equation second-order quantities (involving the product of the two first-order quantities dT and v_n in their notation) are used. An alternative simplified theory is therefore given below, and a more detailed discussion of the problem in § 9. We consider that the treatment given by Meyer and Band (1947) is illogical, as they commence by writing down equations appropriate to a single fluid, and finally divide it into two effective parts.

We may form the following physical picture of the wave motion associated with second sound: due to the relative motion of normal fluid which carries entropy, and superfluid which does not, we obtain a state of the liquid such that alternate layers contain an excess or a deficit of entropy. It follows that these layers are at an effectively higher and lower temperature respectively. In the following simplified deduction of the velocity of such waves, we shall neglect the correction caused by the small pressure differences which will be set up in the liquid due to the fact that the layers at the higher temperatures will contract slightly, and the layers at the lower temperature will expand a little, helium II having a negative coefficient of thermal expansion. We shall also neglect the difference between total and partial differentiation with respect to time, since this difference is quadratic in the velocities—that is, we neglect the difference between differentiation with respect to time whilst remaining stationary in space, and differentiation whilst following the movements of the particles.

Let T be the temperature excess over the mean undisturbed value, ρ the density, S the entropy per unit mass of the liquid considered as a whole, ρ_n the density of normal fluid, ρ_s the density of superfluid, t the time, V_n the velocity of the normal fluid, V_s the velocity of the superfluid, and ϕ the thermodynamic potential of the fluid per unit mass.

Then $\rho=\rho_n+\rho_s$, and for the current j we have $j=\rho_n V_n+\rho_s V_s$. If we assume for the present that entropy is carried only by the normal fluid (we return to this question in § 9), we have the equation of conservation of entropy,

$$\partial(\rho S)/\partial t + S \nabla \cdot (\rho V_n) = 0.$$

Since $C = T' \partial S / \partial T$, this equation may be rewritten (provided the fluctuations are small and density changes negligible)

$$\partial T / \partial t + (T' S / C) \nabla \cdot V_n = 0. \quad \dots\dots(1)$$

Here T' is the actual temperature of the fluid.

Confining attention to waves in which the fluctuations in pressure and density are negligible (these pressure fluctuations being associated with ordinary first

sound), we see from the conservation equation $\partial\rho/\partial t + \nabla \cdot j = 0$ that $j = 0$, as any other constant would imply a constant net flow of the fluid in one direction.

The final equation required is provided by an argument similar to that given by Landau (1941), deducing the equation giving the relation between force and acceleration for the superfluid alone. We imagine two thermally isolated vessels, connected by a capillary so narrow that only superfluid may pass along it. Then the net work required to transfer unit mass of superfluid from one vessel to the other will equal the difference $\partial\phi$ in the thermodynamic potential (Gibbs function) per unit mass between the two vessels. This work will give the change in potential energy. Thus if we consider again the bulk liquid, the gradient of the thermodynamical potential with respect to distance (taken with a negative sign) will give the effective force acting on the superfluid alone. Therefore, for the superfluid, the equation relating acceleration to force is given by $\partial V_s/\partial t + \nabla\phi = 0$. Since $\nabla\phi = -S\nabla T + V\nabla p = -S\nabla T$ for displacements where ∇p is negligible, $\partial V_s/\partial t - S\nabla T = 0$. Equating j to zero, we have $\rho_n V_n = -\rho_s V_s$, and hence

$$\partial V_n/\partial t + S\rho_s \nabla T/\rho_n = 0. \quad \dots\dots(2)$$

Combining this with equation (1), we obtain a wave equation

$$\frac{\partial^2 T}{\partial t^2} = \frac{T'S^2}{C} \frac{\rho_s}{\rho_n} \nabla^2 T, \quad \dots\dots(3)$$

with which there is associated a velocity $u_2 = (T'S^2\rho_s/C\rho_n)^{1/2}$.

We shall now show that in a travelling plane wave of second sound the heat flow at any point is proportional to the temperature at that point, and not to the temperature gradient as in a conducted wave, and also that in a stationary plane wave of second sound the heat flow at any point and at any time is $\frac{1}{2}\pi$ out of phase both in time and in space with the temperature at that point. These results follow from equation (2). The rate of heat flow per unit area is $T'\rho_s V_n$, and thus the rate of heat flow is proportional to the velocity of the normal fluid. In a travelling wave we may take T proportional to $\cos(\omega t - \gamma z)$, where γ is the wave number ω/u_2 . Thus $\partial V_n/\partial t \propto \partial T/\partial z \propto \sin(\omega t - \gamma z)$, and hence $V_n \propto \cos(\omega t - \gamma z) \propto T$. In a stationary wave we may assume T proportional to $\cos \omega t \cos \gamma z$ with suitable origins of time and space. V_n is then proportional to $\sin \omega t \sin \gamma z$. Thus the velocity of the normal fluid, to which the rate of heat flow is proportional, is $\frac{1}{2}\pi$ out of phase with the temperature both in space and in time.

§4. THE ENERGY FLOW IN SOUND

Energy considerations are valuable in the deduction of reflection coefficients and allied problems, and it is therefore important to determine the rate of energy flow per unit area in sound waves.

In a travelling wave of first sound, the kinetic energy per unit volume is given by $\frac{1}{2}\rho V^2$, and if the wave is sinusoidal, the average density of kinetic energy is given by $\frac{1}{4}\rho(V^0)^2$, where V^0 is the peak velocity. It is easily verified that the density of potential energy is the same as that of kinetic energy in a plane wave of first sound (§10); it follows that the total energy density is $\frac{1}{2}\rho(V^0)^2$. Therefore the rate of energy flow of first sound per unit area is given by $\frac{1}{2}\rho(V^0)^2 u_1$, where u_1 is the velocity of the waves. Since in first sound the maximum excess pressure p^0 is given in terms of the maximum velocity V^0 by the equation $p^0 = \rho u_1 V^0$, we may also write for the rate of energy flow $\frac{1}{2}(p^0)^2/\rho u_1$.

In a travelling wave of second sound, the kinetic energy per unit volume is $\frac{1}{2}(\rho_n V_n^2 + \rho_s V_s^2)$. Since $\rho_n V_n = -\rho_s V_s$ this may be rewritten as $\frac{1}{2}\rho_n \rho V_n^2/\rho_s$ or $\frac{1}{2}\rho_n \rho (V_n^0)^2/\rho_s$, where V_n^0 is the peak value of the velocity of the normal fluid. It will be proved later (§ 10) that the density of potential energy has the same value, and hence we may deduce that the total energy density is given by $\frac{1}{2}\rho_n \rho (V_n^0)^2/\rho_s$. Thus the rate of energy flow of second sound per unit area is $\frac{1}{2}\rho_n \rho (V_n^0)^2 u_2/\rho_s$, where u_2 is the velocity of this sound. Since in second sound the maximum excess temperature is given by $T^0 = \rho_n u_2 V_n^0/\rho_s S$, we may also write for the rate of energy flow $\rho \rho_s S^2 (T^0)^2/2\rho_n u_2 = \rho u_2 C (T^0)^2/2T'$, where T' is the mean temperature of the liquid.

§ 5. THE ATTENUATION OF SECOND SOUND DUE TO OTHER HEAT PROCESSES

It is interesting to attempt to calculate the attenuation of second sound caused by non-reversible conduction of heat. For the sake of simplicity only processes in which heat flow is directly proportional to the temperature gradient (with a factor of proportionality K per unit cross-section per unit time) will be considered. Thus included in K will be, for example, the ordinary heat conductivity without fluid movement.

In the equation of entropy conservation we must now consider also irreversible processes. If we consider unit volume, the rate of loss of entropy is $-\partial(\rho S)/\partial t$. This must be equated to the sum of the reversible loss $\rho S \nabla \cdot V_n$ (assuming that only the normal phase carries entropy), and the irreversible loss $(1/T')(\partial Q/\partial t) = -(K/T')\nabla^2 T$, where K is the irreversible conductivity defined above. Thus $\partial(\rho S)/\partial t + \rho S \nabla \cdot V_n - (K/T')\nabla^2 T = 0$ and, therefore,

$$\frac{\partial T}{\partial t} + \frac{T'S}{C} \nabla \cdot V_n - \frac{K}{\rho C} \nabla^2 T = 0.$$

Eliminating V_n between this and equation (2), which is unchanged to this approximation, we find for the (complex) velocity u_2 ,

$$u_2^2 = \frac{T'S^2 \rho_s}{C \rho_n} \left(1 + \frac{K i \omega \rho_n}{\rho T' S^2 \rho_s} \right).$$

Now the amplitude of the wave is proportional to $\exp(i\omega z/u_2)$, of which the real part decays as $\exp(-K\omega^2 z/2u_2^3 \rho C)$. Thus the distance in which the amplitude falls to $1/e$ of its initial value is given by $2u_2^3 \rho C/K\omega^2$. Since $u_2 = 0$ at the λ -point, we see that the attenuation increases rapidly in the region of the λ -point. It also increases near the absolute zero since here C becomes very small, whilst u_2 should remain finite according to theory (§ 11). Thus the attenuation should exhibit a minimum between 0° K. and the λ -point.

For the ordinary non-reversible thermal conductivity K it is reasonable to insert the value found for helium I, where it is not masked by second-sound convection effects. Keesom and Keesom (1936) give $K \sim 6 \times 10^{-5} \text{ cal/deg.cm.sec.}$ at 3.3° K. For the specific heat we insert the smoothed values given by Keesom (1942). Then very approximately the distance in which the amplitude falls to $1/e$ of its initial value is $2 \times 10^{30}/f^2 \text{ cm.}$ at 1.6° K. and $7 \times 10^{10}/f^2 \text{ cm.}$ at 2.18° K. , falling rapidly to zero at 2.19° K. , the λ -point. It therefore appears likely that the attenuation due to irreversible heat effects is only of importance at very high frequencies except close to the λ -point and probably, also, at extremely low temperatures. In the above, f is the frequency in cycles/sec.

§ 6. THE REFLECTION OF SECOND SOUND AT WALLS HAVING FINITE HEAT CONDUCTIVITY

The solution of this problem of reflection is important because one experimental method for measuring the velocity of second sound depends on the setting up of standing waves in a tube. Suppose that there is a source of temperature waves at one end of the tube, and a solid block of material (e.g. a thermometer) of thermal conductivity K at the other. As the heat conductivity, or rather heat convection, of helium II is so much greater than that of any solid, we shall assume that the solid block is effectively infinite in depth. In analogy with transmission-line problems we consider the problem of reflection at a junction of two "transmission tubes", the liquid being one and the solid the other. Let the temperature T be the counterpart of voltage, and the heat flux I the counterpart of current. The ratio T/I across a junction will be called the impedance Z across that junction. Then the problem may be divided into three parts: first, the calculation of the characteristic impedance Z_0 of the liquid; second, the calculation of the impedance Z_l of the solid at the interface between solid and liquid; and finally, the deduction of the reflection and transmission coefficients from these impedances.

For simple harmonic waves of frequency $\omega/2\pi$, equations (1) and (2) of § 3 may be written $\partial I/\partial x = -i\omega C_p T$ and $\partial T/\partial x = -i\omega I/\rho C u_2^2$, where I is the heat flux, $\rho S T V_n$.

Now for a loss-free transmission line of inductance L^* per unit length and capacity C^* per unit length, we have as equations linking voltage V and current I

$$\partial I/\partial x = -i\omega C^* V \quad \dots\dots (4)$$

and

$$\partial V/\partial x = -i\omega L^* I. \quad \dots\dots (5)$$

Comparing these equations with those for our case, we may put $C^* = C\rho$ and $L^* = 1/\rho C u_2^2$. (From § 5 we may also derive the value $G = \omega^2 K_{\text{liq}}/u_2^2$ for the effective resistance per unit length across the tube.) We may now take over the well known theory of loss-free transmission lines. For instance, we may write the characteristic impedance of the liquid as $Z_0 = \sqrt{(L^*/C^*)} = 1/\rho C u_2$. The temperature $T(x)$ at a point x and the heat flux $I(x)$ are given by

$$T(x) = T_l e^{-i\beta x} + T_r e^{i\beta x}, \quad \dots\dots (6)$$

$$Z_0 I(x) = T_l e^{-i\beta x} - T_r e^{i\beta x}, \quad \dots\dots (7)$$

where T_l and T_r are constants giving the temperature maxima of the forward and reflected waves respectively, and β is the wave number ω/u_2 . These formulae satisfy equations (4) and (5).

The standing wave ratio (i.e. the ratio of maximum to minimum) will be given by $(|T_l| + |T_r|)/(|T_l| - |T_r|)$. Using (6) and (7) to determine this in terms of Z_0 and Z_l (the impedance at the end $x=l$), we have for the standing wave ratio $|Z_l/Z_0|$. By a double application of (6) and (7) we may also show that the ratio of temperature to heat flux at a distance l from a junction, across which there is an impedance Z_l , is given by

$$\frac{T}{I} = Z_0 \left\{ \frac{Z_l \cos \beta l + i Z_0 \sin \beta l}{i Z_l \sin \beta l + Z_0 \cos \beta l} \right\}.$$

Next we calculate Z_l , the impedance of the solid at the interface. It follows from § 2 that the temperature in the solid at a depth z from the plane of contact is given by

$$T_{\text{sol}} = T_0 e^{i\omega t} \exp \{ -z\gamma(1+i) \},$$

where $\gamma^2 = \omega \rho_{\text{sol}} C_{\text{sol}} / 2K$, ρ_{sol} and C_{sol} being respectively the density and the specific heat of the solid. The rate of flow of heat into the block per unit area is then given by $I_{\text{sol}} = -K(\partial T_{\text{sol}} / \partial z)_{z=0} = \alpha(1+i)T_0 e^{i\omega t}$, where $\alpha = K\gamma = (\frac{1}{2}\omega \rho_{\text{sol}} C_{\text{sol}} K)^{\frac{1}{2}}$. Thus the impedance at the interface is given by $Z_i = 1/\alpha(1+i)$. The standing wave ratio is $|Z_i/Z_0| = u_2 \rho C / \alpha \sqrt{2}$.

Finally, we must find the reflection and transmission coefficients in terms of Z_0 and Z_i . These may be obtained with the aid of the two boundary conditions that the temperature must be continuous across the interface, and that the heat flow must be a continuous function of distance.

Taking the origin of coordinates at the interface, the first boundary condition gives $T_i + T_r = T_{\text{sol}}$, and the second $I_i + I_r = I_{\text{sol}}$. Using the relations $T_i = Z_0 I_i$, $T_r = -Z_0 I_r$, and $T_{\text{sol}} = Z_i I_{\text{sol}}$, we obtain a reflection coefficient

$$I_r/I_i = (Z_i - Z_0)/(Z_i + Z_0),$$

and a transmission coefficient

$$I_{\text{sol}}/I_i = 2Z_0/(Z_i + Z_0) \simeq 2Z_0/Z_i.$$

The (complex) transmission coefficient is thus given by $2\alpha(1+i)/\rho C u_2$.

§ 7. THE REFLECTION OF SECOND SOUND AT A LIQUID-VAPOUR INTERFACE

In an experimental arrangement originally suggested by Onsager, and used in the experiments of Lane and co-workers (1947), stationary plane waves of second sound are set up in a tube of liquid helium II and, owing to imperfect reflection at the liquid-vapour interface at one end, first sound is transmitted in the vapour. The process may be pictured very simply as follows: the stationary waves of second sound in the liquid cause a periodic fluctuation in temperature at the surface of the liquid, since, as there is very little heat flow at the interface, this is a temperature antinode. There is, therefore, a periodic fluctuation in the amount of vapour emitted from the surface; this produces first sound in the vapour. In this section we demonstrate how the reflection coefficient for such a process may be calculated.

Suppose that a travelling wave of second sound of unit amplitude is incident on the liquid-vapour interface, and that a wave of amplitude R is reflected, the remainder being transmitted into the vapour as first sound. The boundary conditions which will be assumed are: (i) the vapour pressure at the interface is the equilibrium pressure at the temperature of the surface of the liquid; (ii) the difference in the energies transmitted by the incident and reflected waves in the liquid is equal to the energy transmitted by the sound wave in the vapour.

In the liquid the temperature at any point is given by $T_{\text{liq}} = T_0 e^{i\omega t} (e^{-ikx} + R e^{ikx})$. Here $\omega/2\pi$ is the frequency of the wave, k the wave number ω/u_2 , and R is the reflection coefficient.

At the interface we have $T_{\text{liq}} = T_0(1+R)$. Now from the first boundary condition we may deduce that the excess pressure in the first sound wave at the interface is given by $p_0 = T_0(1+R) \partial p / \partial T$, where $\partial p / \partial T$ is the slope of the equilibrium vapour pressure with respect to temperature; this is given by the Clausius-Clapeyron equation $\partial p / \partial T = L / (V_{\text{vap}} - V_{\text{liq}})$.

The excess pressure at any point in the vapour is given by $p = p_0 \exp(i\omega t - ikx)$. Using the results of § 4 we deduce that the rate of energy flow per unit area in the

vapour is $\frac{1}{2}\rho_0^2/\rho_v u_1$, and the difference in rates between the energy flow in the incident and reflected waves in the liquid is $\frac{1}{2}\rho_1 u_2 C T_0^2(1-R^2)/T'$. From equation (5)

$$\left(\frac{\partial p}{\partial T}\right)^2 \frac{(1+R)^2}{\rho_v u_1} = \frac{\rho_1 u_2 C}{T'} (1-R^2),$$

whence
$$R = \left(1 - \left(\frac{\partial p}{\partial T}\right)^2 \frac{T'}{\rho_1 \rho_v u_1 u_2 C}\right) / \left(1 + \left(\frac{\partial p}{\partial T}\right)^2 \frac{T'}{\rho_1 \rho_v u_1 u_2 C}\right).$$

§8. THE PROPAGATION OF SECOND SOUND IN THIN TUBES

For convenience in experimental technique, the investigation of second sound is carried out in tubes of diameter about 1.5 cm. It is therefore important to enquire whether the velocity is appreciably affected by the narrowness of the tube. The physical reason of the expected effect is that the normal fluid has a viscosity and, therefore, sticks to the walls of the tube, whilst the superfluid has no viscosity and, therefore, does not stick.

Consider a plane wave propagated down a circular tube of radius r with viscosity of the normal part of the fluid μ . Then the force acting on the superfluid is that due to the gradient of the thermodynamic potential per unit mass, assuming the entropy to be zero, and the force acting on the normal fluid is that due to the sticking effect. The mass-acceleration of the fluid as a whole will be equal to the sum of the sticking force and the gradient of pressure.

By an argument similar to that given by Rayleigh (1926), we obtain the equation

$$\frac{\partial^2 \rho}{\partial t^2} \left(1 + \frac{2}{r} \sqrt{\frac{\mu}{2\omega\rho}}\right) + \frac{2}{r} \frac{\partial \rho}{\partial t} \sqrt{\frac{\mu\omega}{\rho}} = \nabla^2 p.$$

This equation should hold provided the tube is sufficiently wide for the thickness adhering to the walls to be small compared with the diameter of the tube. Provided the correction terms we are calculating are small, we may rewrite this equation as $\partial^2 \rho / \partial t^2 = (1/f) \nabla^2 p$, where

$$f = 1 + (2/r) \sqrt{(\mu/2\omega\rho)} + (2/i\omega r) \sqrt{(\mu\omega/\rho)}.$$

This equation, with $\partial^2 S / \partial t^2 = S^2 \rho_s \nabla^2 T / \rho_n$ †, gives the velocity of the two sounds. By a slight extension of the work of Landau (1941), we obtain

$$u^4 - u^2 \left\{ \frac{1}{f} \left(\frac{\partial p}{\partial \rho} \right)_s + \frac{T' S^2 \rho_s}{C \rho_n} \right\} + \frac{T' S^2 \rho_s}{C \rho_n f} \frac{\partial p}{\partial \rho} = 0.$$

According to this equation, the velocity of first sound is given by

$$u_1 = \sqrt{\left(\frac{1}{f} \frac{\partial p}{\partial \rho} \right)} \simeq \frac{\partial p}{\partial \rho} \left(1 - \frac{1}{r} \sqrt{\frac{\mu}{2\omega\rho}} \right),$$

and the velocity of second sound by $u_2 = (T' S^2 \rho_s / C \rho_n)^{\frac{1}{2}}$. Thus the correction to the velocity of second sound, on account of the narrowness of the tube, is much smaller than that to the velocity of first sound. ‡

† This equation (cf. (3) § 3) is true even in the case where pressure differences are not negligible (Landau 1941).

‡ Note added in proof. A more detailed consideration now shows that there is a very small correction proportional to the product of the coefficient of thermal expansion and the square root of the viscosity.

§9. THE USE OF THE LAGRANGIAN AND THE TOTAL ENERGY IN THE DEDUCTION OF THE VELOCITY OF SECOND SOUND IN A FLUID CONSISTING OF PARTS WITH TWO DIFFERENT ENTROPIES

There is some question as to whether the entropy of the superfluid in helium II is really exactly zero; we shall therefore discuss and apply a method which is an extension of that of Tisza (1947).

Let the entropy of the normal fluid alone be S_n , and the entropy of the superfluid is S_s . Then the measured entropy S of the whole fluid is given by $\rho S = \rho_n S_n + \rho_s S_s$. Now we introduce new coordinates ξ and η such that $\rho\xi = \rho_n V_n + \rho_s V_s$ and $\eta = (\rho_n V_n S_n + \rho_s V_s S_s) / (\rho_n S_n + \rho_s S_s) - \xi$. Since η is defined as the excess over ξ , the velocity of the fluid as a whole, of the velocities of the normal and the superfluid, weighted according to the amount of entropy which they carry, the equation of the conservation of entropy is expressible in these new coordinates in the simple form $\partial S / \partial t + S \nabla \cdot \eta = 0$. Also in these coordinates, the equation of the conservation of mass becomes $(1/\rho)(\partial \rho / \partial t) + \nabla \cdot \xi = 0$. After some reduction, using the fact that $\rho = \rho_n + \rho_s$, we obtain for η , $S^2 \rho^2 \eta = \rho_n \rho_s S^* (V_n - V_s)$, where $S^* = S_n - S_s$. Thus the transfer of entropy depends on the relative velocities of the parts with different entropies, as is clear physically. From these equations we may obtain V_n and V_s in terms of ξ and η as follows:

$$V_n = \xi + (\eta \rho S / \rho_n S^*); \quad V_s = \xi - (\eta \rho S / \rho_s S^*).$$

Hence the kinetic energy per unit volume is given by

$$\frac{1}{2} \rho_n V_n^2 + \frac{1}{2} \rho_s V_s^2 = \frac{1}{2} \rho \xi^2 + \frac{1}{2} (\rho^3 S^2 \eta^2 / \rho_n \rho_s S^{*2}).$$

The potential energy is due to fluctuations both in volume and in entropy. Now if δU represents a small change in internal energy, we may expand in terms of δV , the small change in volume, and δS , the small change in entropy. Retaining terms up to the second order, we have

$$\begin{aligned} \delta U &= \left(\frac{\partial U}{\partial V} \right)_S \delta V + \left(\frac{\partial U}{\partial S} \right)_V \delta S + \frac{1}{2} \frac{\partial^2 U}{\partial V^2} \delta V^2 + \frac{\partial^2 U}{\partial V \partial S} \delta V \delta S + \frac{1}{2} \frac{\partial^2 U}{\partial S^2} \delta S^2 \\ &= -p \delta V + T \delta S - \frac{1}{2} \left(\frac{\partial p}{\partial V} \right)_S \delta V^2 - \left(\frac{\partial p}{\partial S} \right)_V \delta V \delta S + \frac{1}{2} \left(\frac{\partial T}{\partial S} \right)_V \delta S^2. \end{aligned}$$

Thus the minimum work to cause the fluctuations is given by (Landau and Lifshitz 1938)

$$\begin{aligned} \delta U + p \delta V - T \delta S &= -\frac{1}{2} \left(\frac{\partial p}{\partial V} \right)_S \delta V^2 - \left(\frac{\partial p}{\partial S} \right)_V \delta V \delta S + \frac{1}{2} \left(\frac{\partial T}{\partial S} \right)_V \delta S^2 \\ &= -\frac{1}{2} \left(\frac{\partial p}{\partial V} \right)_S \delta V^2 - \left(\frac{\partial p}{\partial T} \right)_V \left(\frac{\partial T}{\partial S} \right)_V \delta V \delta S + \frac{T'}{2C} \delta S^2 \\ &= \frac{1}{2} \left(\frac{\partial p}{\partial \rho} \right)_S \left(\frac{\delta \rho}{\rho} \right)^2 - \rho \alpha \left(\frac{\partial p}{\partial \rho} \right)_T \frac{T'}{C} \delta V \delta S + \frac{T' S^2}{2C} \left(\frac{\delta S}{S} \right)^2. \end{aligned}$$

Here α is the coefficient of thermal expansion, which is small, and may be neglected; the inclusion of this term gives the coupling between first and second sound. Thus the potential energy is, per unit mass,

$$\frac{1}{2} \left(\frac{\partial p}{\partial \rho} \right)_S \left(\frac{\delta \rho}{\rho} \right)^2 + \frac{T' S^2}{2C} \left(\frac{\delta S}{S} \right)^2.$$

It is convenient to rewrite the equations of continuity and of entropy conservation in terms of displacements rather than velocities. If we suppose that ξ_1 is the displacement corresponding to the velocity ξ , and η_1 is the displacement corresponding to the velocity η , we may write the equations of continuity and of entropy conservation as

$$\delta\rho/\rho = -\nabla \cdot \xi_1 \quad \text{and} \quad \delta S/S = -\nabla \cdot \eta_1.$$

The potential energy per unit volume becomes in these coordinates

$$\frac{\rho}{2} \left(\frac{\partial p}{\partial \rho} \right)_S (\nabla \cdot \xi_1)^2 + \frac{\rho T' S^2}{2C} (\nabla \cdot \eta_1)^2,$$

and the kinetic energy per unit volume becomes

$$\frac{1}{2} \rho \dot{\xi}_1^2 + \frac{1}{2} \frac{\rho^3}{\rho_n \rho_s} \left(\frac{S}{S^*} \right)^2 \dot{\eta}_1^2,$$

where dots denote differentiation with respect to time. Thus the total energy density is

$$\frac{1}{2} \rho \dot{\xi}_1^2 + \frac{1}{2} \frac{\rho^3}{\rho_n \rho_s} \left(\frac{S}{S^*} \right)^2 \dot{\eta}_1^2 + \frac{1}{2} \left(\frac{\partial p}{\partial \rho} \right)_S (\nabla \cdot \xi_1)^2 + \frac{\rho T' S^2}{2C} (\nabla \cdot \eta_1)^2,$$

and the Lagrangian density is

$$\frac{1}{2} \rho \dot{\xi}_1^2 + \frac{1}{2} \frac{\rho^3}{\rho_n \rho_s} \left(\frac{S}{S^*} \right)^2 \dot{\eta}_1^2 - \frac{1}{2} \left(\frac{\partial p}{\partial \rho} \right)_S (\nabla \cdot \xi_1)^2 - \frac{\rho T' S^2}{2C} (\nabla \cdot \eta_1)^2.$$

The appropriate Lagrangian equations of motion are

$$\frac{d}{dt} \left(\frac{\partial L}{\partial \dot{\xi}_1} \right) - \frac{\partial L}{\partial \xi_1} = -\nabla \cdot \frac{\partial L}{\partial (\nabla \cdot \xi_1)} \quad \text{and} \quad \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{\eta}_1} \right) - \frac{\partial L}{\partial \eta_1} = -\nabla \cdot \frac{\partial L}{\partial (\nabla \cdot \eta_1)}.$$

Here $\partial L / \partial (\nabla \cdot \xi_1)$ and $\partial L / \partial (\nabla \cdot \eta_1)$ may be interpreted as pressures, and the gradients of these pressures give rise to forces. We obtain then the wave-equations $\ddot{\xi}_1 = (\partial p / \partial \rho)_S \nabla^2 \xi_1$ and $(\rho^2 / \rho_n \rho_s) (S / S^*)^2 \ddot{\eta}_1 = (T' S^2 / C) \nabla^2 \eta_1$.

The velocity of the first sound associated with the first of these equations is, therefore, $u_1 = (\partial p / \partial \rho)^{1/2}$, and the velocity of second sound as given by the second is $u_2 = (T' \rho_n \rho_s S^{*2} / C \rho^2)^{1/2}$.

We must again emphasize at this point that $S^* = S_n - S_s$ is the difference in entropies between the normal and the superfluid when account is taken of the fact that the fraction of the normal fluid is ρ_n / ρ and the fraction of superfluid is ρ_s / ρ . If we could isolate the superfluid, and it had the same density ρ as the bulk fluid, we could measure its entropy and call it, say, S'_s . Similarly, if we had only normal fluid of density ρ , its entropy might be called S'_n . We should then have $\rho_s S_s = \rho S'_s$ and $\rho_n S_n = \rho S'_n$. Then

$$S^* = S_n - S_s = \rho (S'_n / \rho_n - S'_s / \rho_s).$$

In this notation

$$u_2^2 = (T' / C) \{ \rho_s (S'_n)^2 / \rho_n + \rho_n (S'_s)^2 / \rho_s - 2 S'_n S'_s \}.$$

This formula reduces to the well known one $u_2^2 = T' S^2 \rho_s / C \rho_n$ when any entropy associated with the superfluid is ignored.

§ 10. THE AVERAGE EQUALITY OF POTENTIAL AND KINETIC ENERGY IN A PLANE WAVE OF SECOND SOUND

We may now prove the proposition of § 4 that the time averages of the potential and kinetic energies are equal. This proposition may be stated in the alternative form of the vanishing of the time-average of the Lagrangian.

The Lagrangian density is, as shown in the last section,

$$\left[\frac{1}{2} \rho \dot{\xi}_1^2 - \frac{1}{2} \rho \left(\frac{\partial \rho}{\partial \rho} \right)_S (\nabla \cdot \xi_1)^2 \right] + \left[\frac{1}{2} \frac{\rho^3}{\rho_n \rho_s} \left(\frac{S}{S^*} \right)^2 \dot{\eta}^2 - \frac{1}{2} \frac{T' S^2 \rho}{C} (\nabla \cdot \eta_1)^2 \right],$$

the terms in the first bracket being contributed by first sound and those in the second by second sound. If we suppose that ξ_1 and η_1 are displacements in plane sine waves, we shall have

$$\xi_1 = \xi_1^0 \sin \omega(t - x/u_1) \quad \text{and} \quad \eta_1 = \eta_1^0 \sin \omega(t - x/u_2),$$

and thus

$$\dot{\xi}_1^2 = \frac{1}{2} (\omega \xi_1^0)^2 \quad \text{and} \quad \dot{\eta}_1^2 = \frac{1}{2} (\omega \eta_1^0)^2.$$

Also

$$(\partial \xi_1 / \partial x)^2 = \frac{1}{2} (\omega \xi_1^0 / u_1)^2 = \frac{\dot{\xi}_1^2}{u_1^2}$$

and

$$(\partial \eta_1 / \partial x)^2 = \frac{1}{2} (\omega \eta_1^0 / u_2)^2 = \frac{\dot{\eta}_1^2}{u_2^2}.$$

Using the values deduced for u_1 and u_2 in the last section, we see at once that each of the two brackets forming the Lagrangian density vanishes when averaged over time. Since by Fourier analysis any plane wave may be split into a series of pure sine waves, this result holds for all plane waves. The result is thus a special case of the general proposition of the extremum of the time-integral of the Lagrangian function.

§ 11. THE SHAPE OF THE VELOCITY-TEMPERATURE CURVE FOR SECOND SOUND ON THE BASIS OF LANDAU'S THEORY

Landau (1941) used the formula $u_2^2 = TS^2 \rho_s / C \rho_n$ for the velocity of second sound with calculated as opposed to experimental (or empirical) values of the thermodynamical quantities involved. These calculated expressions contain at least two adjustable constants (arising in the roton contributions in Landau's terminology: in the case of a sharp energy gap as considered by Landau in his 1941 paper, just two constants, Δ and μ , are required, Δ giving the magnitude of the energy gap and μ the effective mass of the roton). The statement made originally by Lifshitz (1944) that Landau's expressions give for the velocity of second sound a monotonic decreasing function of temperature is only true for certain ranges of values of Δ and μ . It will now be shown that if the rôle of phonons is ignored (using Landau's terminology: phonons are Debye waves), the velocity at about 1.5° K. should rise with temperature, and this value should be almost independent of the particular value of Δ chosen. It is therefore possible to choose μ and Δ in such a way that the roton contributions overweigh those of the phonon contributions and yet still give the correct magnitude for the velocity of the sound. In this way it is possible to obtain a dependence of velocity of second sound on temperature far from monotonic, having a new maximum between the λ -point and the maximum at absolute zero. It is impossible as yet to deduce Δ except crudely, as inertia (Andronikashvili 1946) and specific heat experiments are not sufficiently accurate; as has been pointed out already, the velocity of second sound is largely independent of Δ in the range of temperatures so far employed for these experiments.

To deduce this from Landau's theory (1941), we start with the formula $u_2^2 = TS^2 \rho_s / C \rho_n$. Now at about 1.5° K. $\rho_s \simeq \rho$, the factor ρ_s / ρ_n merely giving the sharp decrease of velocity near the λ -point. From Landau's paper we may deduce

$$(\rho_n)_{\text{roton}} \propto u^{5/2} T^{3/2} e^{-\Delta/KT},$$

$$(C)_{\text{roton}} \propto n^{3/2} T^{-1/2} \Delta^2 e^{-\Delta/KT} \left\{ 1 + \frac{3KT}{\Delta} + \frac{15}{4} \left(\frac{KT}{\Delta} \right)^2 \right\},$$

$$(S)_{\text{roton}} \propto u^{3/2} T^{1/2} \Delta e^{-\Delta/KT} \{ 1 + 5KT/2\Delta \}.$$

Thus, when only roton contributions are of importance,

$$u_2^2 = \frac{TS^2\rho_s}{C\rho_n} \propto \frac{T}{\mu} \left\{ 1 + \frac{5KT}{2\Delta} \right\} / \left\{ 1 + \frac{3KT}{\Delta} + \frac{15}{4} \left(\frac{KT}{\Delta} \right)^2 \right\}.$$

Since $KT \ll \Delta$ at $T \sim 1.5^\circ \text{K.}$, we see that under these conditions u_2 must rise with temperature.

Landau's results that the maximum value of the velocity is at absolute zero, and is equal to $u_1/\sqrt{3}$, will hold provided (i) any roton contributions have vanished due to the energy gap in the roton spectrum, and (ii) one part of the liquid has entropy zero and the other has an entropy due to Debye waves (phonons).

The temperature, according to Landau's theory, at which the rapid rise in the velocity of second sound takes place (at these temperatures close to the absolute zero) is governed chiefly by the value of Δ , which, as we have seen, is very difficult to derive experimentally from other sources. A value as low as about half a degree would be quite a possibility without contradicting present experimental evidence. Even this assumes the existence of a well-defined energy gap Δ . In this latter connection, a recent note by Landau (1947) is interesting.

§ 12. A CRITICISM OF TISZA'S THEORY

We only consider those aspects of Tisza's paper (1947) which directly concern the problem of the propagation of second sound.

Firstly, there is the problem as to whether it is possible for Debye waves to pass through the superfluid. Unlike Landau (1941), who includes phonons (i.e. Debye effects) as part of the normal fluid, Tisza considers that Debye waves are associated with the liquid as a whole, and that their presence will cause an attenuation of second sound. However, in his quantitative theory Tisza neglects their effect. If Tisza's viewpoint is correct, his theory must break down at temperatures below about 1°K. , and his derivation of the zero value of the velocity of second sound at absolute zero (p. 852) and equations (11) and (12)— $d\rho_n/\rho_n = dS_n/S_n$; $\rho_n/\rho = S_n/S_0$ —appear therefore to be invalid. Further, in the deduction of these equations Tisza considers the entropy as trapped in a cell possessing a membrane permeable only to superfluid. The argument of p. 845 is only true provided no heat is supplied from outside, since any heat supplied would create more normal fluid and, therefore, more entropy in the cell. Under the adiabatic conditions the equations are thus correct, but equation (12) will merely tell us the temperature in the cell, and will not, as Tisza claims, apply to equilibrium values at any fixed temperature we like to name. Hence we do not agree that Tisza's equation (13) $\{\rho_n/\rho = (T/T_0)^r\}$, where T_0 is the temperature of the λ -point} follows from (8) $\{S_n/S_0 = (T/T_0)^r\}$, assuming that S_n and ρ_n refer to equilibrium values at any given temperature. Also it is not clear physically why there should exist any thermodynamic relationship between the equilibrium values of S_n and ρ_n for any given temperature; one would expect the actual relationship between them to be specific to the detailed mechanism involved.

The values which are required in the formula for the velocity of second sound are ordinary equilibrium values of S , C and ρ_n for a given temperature T' . These mean equilibrium values,† about which the adiabatic fluctuations comprising second sound occur, are not directly connected with the values relating to Tisza's

† Actually not all the variables in this formula for the velocity of second sound are independent. Since $C = T'(\partial S/\partial T)$ and $\rho_s = \rho - \rho_n$, we may transform the formula to become $u_2^2 = (1 - \rho/\rho_n)/\partial(1/S)/\partial T$. Thus the mean equilibrium values required for any given temperature are those of ρ_n and $\partial(1/S)/\partial T$.

"isopycnic" compression experiment. Thus it is incorrect to apply equation (12) (as Tisza does) to obtain the value of the velocity of second sound at $T=0$, and it is also irrelevant that Landau's phonon equations violate this equation, as Tisza points out on p. 852 of his paper. It must be added that the relationships derived by Landau (1941) for his roton effects also conflict with Tisza's equation (12), as may be seen from the values given in § 11 of this paper. (The entropy and normal density of an ideal Bose-Einstein gas do happen to obey Tisza's relationship, but it is well known that the values calculated on this very crude model are quite inaccurate, the entropy being given as proportional to $T^{1.5}$, as opposed to the experimental relation $S \propto T^{5.5}$.)

The present situation with respect to the temperature variation of the velocity of second sound might be summed up by saying that Landau and Tisza now agree on the formula $u_2^2 = TS^2\rho_n/C\rho_n$, although Tisza believes that the value of S inserted in this formula should refer not to the total entropy, but to this value less the Debye contribution. Tisza also considers that the formula will become inadequate below about 1°K. , where these Debye effects become important. This author relies on empirical formulae for the entropy, with the constants determined from Kapitza's experimental values (Kapitza 1941): he derives his values of ρ_n from this entropy formula by the application of his relationship $\rho_n/\rho = S_n/S_0$, which, as we have shown above, has not the thermodynamic justification he has claimed. Landau goes further and attempts a theory involving two parameters, Δ and μ , capable of giving the values of S , C and ρ_n required, Δ and μ being determined from Keesom's values of the specific heat (Keesom 1942).

ACKNOWLEDGMENTS

The author would like to record his acknowledgment of the many helpful discussions he has had with Professor N. F. Mott of this Laboratory, Professor J. F. Allen of the University of St. Andrews, and colleagues of the Royal Society Mond Laboratory, Cambridge. He is also indebted to the Department of Scientific and Industrial Research for a grant.

REFERENCES

- ANDRONIKASHVILI, 1946, *J. Phys. U.S.S.R.*, **10**, 201.
 GOGATE, D. V., and PATHAK, P. D., 1947, *Proc. Phys. Soc.*, **59**, 457.
 KAPITZA, P., 1941, *J. Phys. U.S.S.R.*, **4**, 181; **5**, 59.
 KEESOM, W. H., 1942, *Helium* (Elsevier Publishing Co.), p. 217.
 KEESOM, W. H., and KEESOM, A. P., 1936, *Physica*, **3**, 359.
 LANDAU, L., 1941, *J. Phys. U.S.S.R.*, **5**, 71; 1944, *Ibid.*, **8**, 1; 1947, *Ibid.*, **11**, 91.
 LANDAU, L., and LIFSHITZ, E., 1938, *Statistical Physics* (Oxford: University Press), p. 114.
 LANE, C. T., *et al.*, 1947, *Phys. Rev.*, **71**, 600.
 LIFSHITZ, E., 1944, *J. Phys. U.S.S.R.*, **8**, 110.
 LONDON, F., 1938, *Phys. Rev.*, **54**, 947; 1939, *J. Phys. Chem.*, **43**, 49.
 MEYER, L., and BAND, W., 1947, *Phys. Rev.*, **71**, 828.
 PESHKOV, V. L., 1946, *J. Phys. U.S.S.R.*, **10**, 389.
 RAYLEIGH, Lord, 1926, *Theory of Sound*, 2nd Edn. (London: Macmillan), p. 318.
 TISZA, L., 1938, *C. R. Acad. Sci., Paris*, **207**, 1035, 1186; 1940, *J. Phys. Radium*, **1**, 165, 350; 1947, *Phys. Rev.*, **72**, 838.

Electronic Structure of Fulvene

By C. A. COULSON*, D. P. CRAIG† AND A. MACCOLL†

* Kings' College, London

† University College, London

MS. received 14 December 1947

IT is a well known fact that in dealing with most common conjugated and aromatic systems, such as benzene or the polyenes, the method of molecular orbitals (m.o.) and the method of valence-bond structures (v.b.) agree remarkably closely. This agreement includes bond orders, bond lengths, free valence and resonance energy. However, during a study of the fulvene molecule $C_5H_4 \cdot CH_2$ we have found a large and serious discrepancy between the predictions of the two methods of approximation. These discrepancies are large enough to be used to distinguish between the reliability of the methods by an appeal to experiment. Unfortunately the structural data are not yet available. But in view of the desirability of obtaining as many comparisons and tests as possible between the theories, it seems good to place on record the calculations that we have made, in the hope that they may stimulate experimental work with which to compare them. The calculations themselves are of quite standard pattern and we shall not therefore describe them in any detail.

In the v.b. calculations, resonance is included between five canonical structures. In the first place these are taken to be A – E.

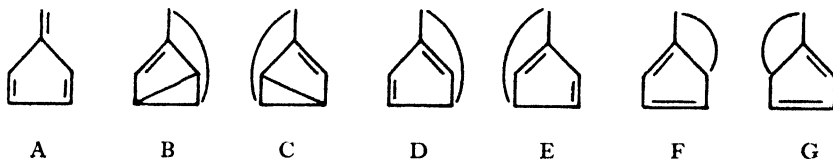


Figure 1.

The normalized wave function for the ground state is

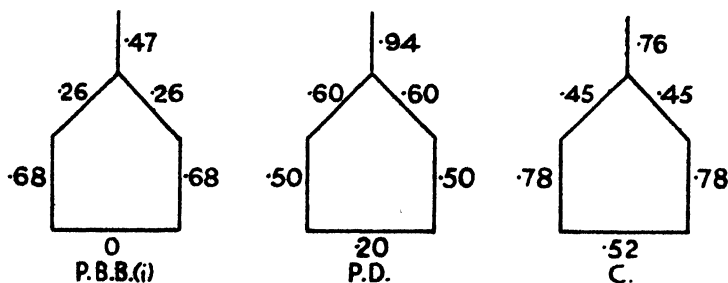
$$\Psi = 0.6324 \psi_A - 0.2109 (\psi_B + \psi_C) + 0.4217 (\psi_D + \psi_E). \quad \dots (1)$$

The resonance energy, 0.5α , has already been given by Wheland (1941). Bond orders may easily be inferred by the method of weighted contributions, as suggested by Pauling, Brockway and Beach (1935), or by the method of Penney (1937), using the Dirac coupling formula. The values are shown in figure 2, with the labels P.B.B. (i) and P.D. respectively.

The m.o. calculation of bond orders are made by the method of Coulson (1939). These values are shown in the figure with the label C. The resonance energy, 1.47β , has already been given by Wheland (1941).

Several comments suggest themselves at once from a study of figure 2. Firstly, the differences in bond order are much greater than is usually found, and even the sequence of orders is not preserved in any two of the diagrams. So far as the P.B.B. calculations are concerned this may be attributed to the impossibility of making any really satisfying choice of canonical structures. An objection

to the choice A-E is that none of them shows the bottom bond as double, so that in the P.B.B. method it cannot have any double bond character. The only way to avoid this is to make a different choice of fundamental structures. Any independent set of five will do, even if, in the language of Pauling, they are not canonical. This suggests that we introduce F and G. The appropriate modification of the wave function (equation (1)) is easily made for any independent



Valence-bond

Molecular-orbital

Figure 2. Calculated bond orders.

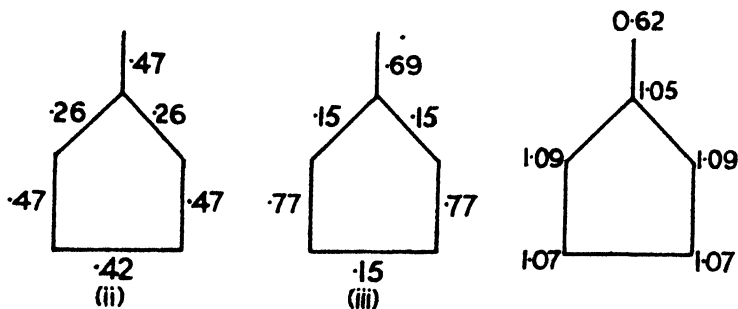


Figure 3. Alternative P.B.B. calculations of bond order.

Figure 4. Charge migration according to m.o. theory. Figures show total mobile charge on each atom.

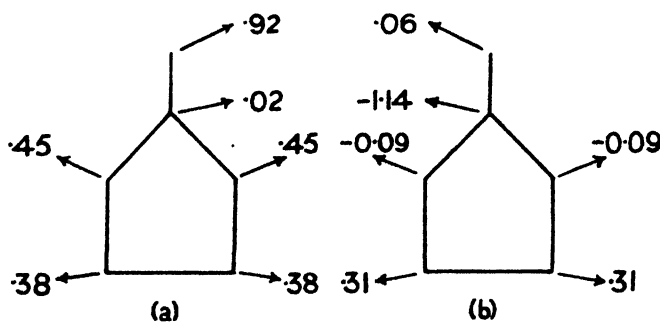


Figure 5. Free valence.

(a) Molecular-orbitals $F.V. = 4.68 - \mathcal{A}^{\circ}$.

(b) Valence-bond, relative to an ethylenic carbon, so that $F.V. = 4 - \mathcal{A}^{\circ}$.

set, using the "un-crossing rule". In this manner we have used three reasonable sets, viz. (i) A-E, (ii) A, B, C, F and G, (iii) A, D, E, F, and G. Of these only (i) is a canonical set, but (iii) has the advantage of not using structures with more than one formal bond. The results of our calculations in case (i) are already given in figure 2: the results for cases (ii) and (iii) are shown in figure 3. Even

the most cursory glance at these diagrams shows conclusively that there is not even the remotest agreement between them. It is quite evident that in molecules of this kind, in which "chemical intuition" does not suggest one single sensible choice of structures, the whole idea of weighted contributions is inadequate. Indeed, it suggests that even in other molecules such as anthracene or chrysene, although the choice of structures is reasonably straightforward, we must treat all calculation of bond orders by the P.B.B. method with some reserve. The same is not true for bond orders calculated according to the P.D. method, for these are invariant for any choice of five independent structures, and we obtain the same values (figure 2) whether we use the set (i), (ii), (iii) or any other independent set of five.

There still remains the large discrepancy between the P.D. valence bond and the m.o. bond orders. It is here that the distinction between the two methods is most clearly seen. For the m.o. method predicts quite large charge shifts. In its conventional form, it discusses the motion of the π electrons in a potential field supposed to be equivalent around each nucleus. But if we calculate the resulting charge distribution we find (figure 4) a large drift of electrons into the five-membered ring. For that reason the potential field around each nucleus cannot be equivalent, so that, as has already been pointed out in general terms (Coulson and Rushbrooke 1940), the m.o. method is not genuinely self-consistent. It is true that the shifts are probably rather less than figure 4 would indicate, but they should be great enough to provide a reasonably large dipole moment. On the other hand there is no dipole moment whatever in the v.b. calculations. Thus, if ionic terms are as important as the m.o. method suggests, then the v.b. method is inadequate in its simple form. But if they are relatively unimportant, then the v.b. method is the more reliable. An experimental determination of the dipole moment would settle this issue.

A further discrepancy between the two methods is found in the calculated energies. For a large number of molecules (Wheland 1941), the ratio of the resonance integral (β) to the exchange integral (α) required to give identical resonance energies has the value $\beta/\alpha = 0.54$. But in the case of fulvene it is 0.34. A really satisfactory heat of combustion or hydrogenation could settle which method was more nearly correct.

Our final comment concerns the "free valence" (F.V.) at any particular centre (Coulson 1948 a). According to the method of m.o., this is given by

$$\text{F.V.} = 4.68 - \mathcal{N}, \quad \dots\dots(2)$$

where $\mathcal{N}$ is the total bond order ($\sigma + \pi$) of all the bonds leaving the given atom. Values obtained in this way are shown in figure 5(a). If we may suppose that a high free valence is indicative of reactivity, then by comparison with benzene, for which the free valence is 0.35, it is seen that the non-ring carbon atom is extremely reactive. Some measure of polymerization, with or without dismutation, would therefore be anticipated. This end carbon atom may be compared with the corresponding atom of benzyl, for which the free valence (Coulson 1948 b) is only slighter greater, at 1.04 instead of 0.92.

In cases where the choice of canonical structures in the resonance method is fairly evident, it is possible (Pullman 1947) to use the method of weighted contributions from each member of the set to estimate free valence, in a manner similar to that previously described for bond orders. But in the case of fulvene,

the reasons which we have given for disregarding the bond orders thus calculated, apply equally to the free valence, and we shall not therefore use this method. A similar objection does not obtain (Moffitt 1948) with the P.D. type of calculation, for which a description of the free valence may be given, parallel to that implied in equation (2) above for the molecular orbital theory. But since the greatest possible value of total bond order at a carbon atom (corresponding to 4.68 in m.o. theory) has not yet been settled for v.b. theory, our best way is to measure free valence relative to an ethylenic carbon. The resulting values are shown in figure 5(b). The high reactivity of the end carbon compared with the other carbons is now not at all noticeable and the difference between figures 5(a) and 5(b) is marked. A more adequate experimental study of the chemistry of fulvene should help to resolve the question which of 5(a) and 5(b) better represents the molecule.

REFERENCES

- COULSON, C. A., 1939, *Proc. Roy. Soc. A*, **169**, 413; 1948 a, *Trans. Faraday Soc.*, **34**, 123; 1948 b, *Ibid.*, "Discussion on Free Radicals".
 COULSON, C. A., and RUSHBROOKE, G. S., 1940, *Proc. Camb. Phil. Soc.*, **36**, 193.
 MOFFITT, W. E., 1948, *Trans. Faraday Soc.*, in the press.
 PAULING, L., BROCKWAY, L. O., and BEACH, J. Y., 1935, *J. Amer. Chem. Soc.*, **57**, 2705.
 PENNEY, W. G., 1937, *Proc. Roy. Soc. A*, **158**, 306.
 PULLMAN, A., 1947, *Ann. Chim.*, **2**, 1.
 WHELAND, G. W., 1941, *J. Amer. Chem. Soc.*, **63**, 2025.

Some Observations concerning the Energy of Nuclei

By E. GLUECKAUF

Atomic Energy Research Establishment, Harwell, Didcot

MS. received 16 December 1947

ABSTRACT. The α -binding energy as a function of the number of neutrons and protons is shown to exhibit sharp fluctuations which occur at more or less regular intervals. There is evidence that these "kinks" are caused by discontinuous changes in the density and radius of the nuclei.

It is shown that a mathematical connection exists between "kinks" in the α -binding energy curve and the irregularities of the β -stability line (Gamow valley), so that both must occur in the same region of the nuclei. This applies also to the irregularities in the mass-defect curve.

A study is made of the energies of β -decay (ΔM) of isobaric nuclei in the region of the radioactive elements. This shows that nuclei of odd mass-number A have a slightly higher binding energy for even proton number Z (a lower mass by 0.2 mev.) than for odd Z . The effect, though well known for nuclei of even mass-numbers, has not before been observed for odd ones, where it is much smaller.

§1. REGULARITIES IN THE α -STABILITY OF NUCLEI

THE purpose of this section is to draw attention to certain regular variations in the α -binding energy of nuclei which are superimposed on the general trend of decreasing α -stability with increasing mass and charge. Being related to the gradient of the mass-defect curve, the α -stability is better suited than the mass defect itself for exhibiting irregularities which point to structural features in nuclei.

We shall employ the usual terminology, denoting the atomic weight by A , the atomic number by Z , the number of neutrons by $N (= A - Z)$, and the isotopic number by $I (= N - Z = A - 2Z)$.

(i) *Data for heavy nuclei*

One of the well known features of all the radioactive series is the extremely marked change in the α -energies between the isotopes of atomic number $Z = 82$,

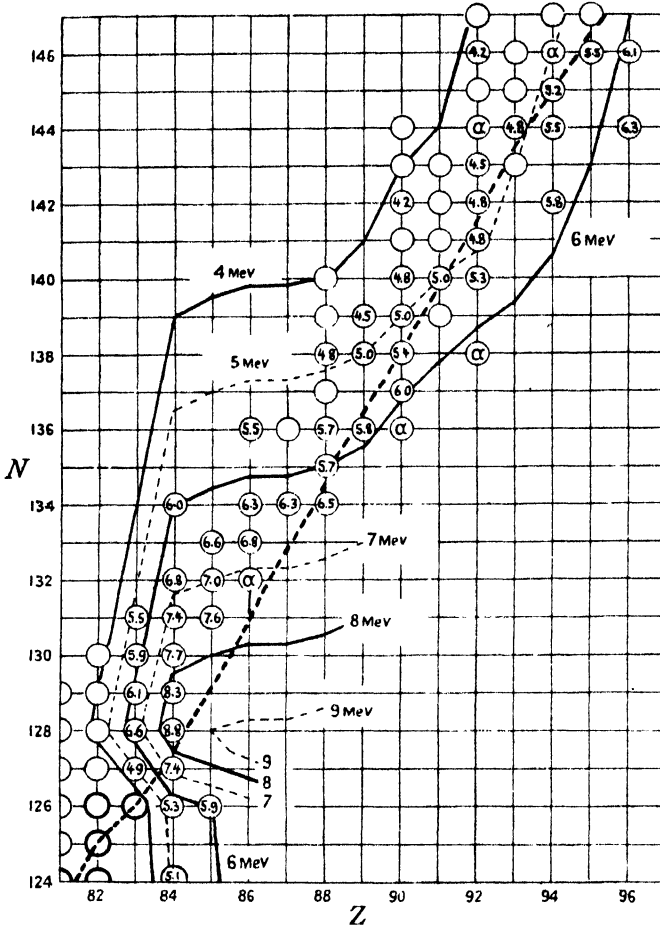


Figure 1. α -energies of large nuclei (mev.).

Heavy circle: Stable nuclei.

Light circle: β -active nuclei.

Light circle surrounding figures: α -active nuclei with or without β -decay.

which are, within the limits of experimental accuracy, completely stable against α -decay, and those of atomic number $Z = 84$, which have the most short-lived bodies and emit the most powerful α -particles. The average rate of change of the α -energy E_α in this region is $(\Delta E_\alpha / \Delta Z)_{N=\text{const.}} = 2.1$ mev. per proton. After element 84, the α -energy increases only slightly, $[\Delta E_\alpha / \Delta Z]_N$, varying from 0.25 to 0 to 0.5 mev. per proton.

This striking discontinuity at the proton number $Z = 84$ has given rise to much valuable discussion (notably by Feather 1945), during which, however,

it has apparently been overlooked that a discontinuity of similar magnitude and importance exists also at the neutron number $N=128$. If we plot the energy of the emerging α -particles against the neutron number N and draw lines of constant Z , we get a sharp rise in E_α between $N=126$ and $N=128$ with a gradient of $[\Delta E_\alpha/\Delta N]_{Z=\text{const.}} = 1.9$ mev. per neutron, followed after $N=128$ by a constant decrease $[\Delta E_\alpha/\Delta N]_Z = -0.5$ mev., which lessens gradually after $N=138$ until it reaches a value of -0.1 mev. at $N=144$.

No systematic diagram of the α -energies is therefore quite complete which does not show this dual system of discontinuities at $Z=84$ and at $N=128$. It would appear that the best way of representing the facts is shown in figure 1. This represents a map of nuclei, using Z and N as abscissa and ordinate respectively, and the α -energy is indicated for every nuclear species. We can then draw contour-lines of constant α -energy, treating that quantity as if it were a continuous variable of Z and N . The diagram is ideally suited for the interpolation and extrapolation of unknown α - and β -energies.

To give an example, it is possible to predict with a fair degree of certainty that the only β -stable α -emitter of element 97 will have mass 247, an E_α of 6.0 to 6.1 mev. and a half-life of the order of one year. As the highest mass number so far reported in a trans-uranic element is 242, it is thus not likely that this new species will be discovered in the near future.

Though in figure 1 the discontinuities at $Z=84$ and $N=128$ are very clearly noticeable, the general impression is that these discontinuities are not as important as they appear if Z or N alone is chosen as variable. It rather appears that the lines of constant E_α form an irregular wave, with its axis somewhat steeper than the band of the known α -emitting nuclei which follow approximately the line of maximum β -stability (dotted line). The question immediately arises whether this E_α "wave" is confined to the radioactive elements or whether it can be traced also in the binding energies of the α -stable nuclei.

(ii) Data for light nuclei

For the lightest elements up to a mass-number A of about 40, the nuclear masses M are well enough known to permit a fairly accurate calculation of the binding energy E'_α for the last α -particle (which is the α -energy with opposite sign), according to

$$E'_\alpha = \frac{A-4}{Z-2}M + \frac{4}{2}M - \frac{A}{Z}M.$$

With reduced accuracy these calculations can be extended to $A=60$. The mass data have been taken from an isotope chart compiled by Perlman, Seaborg and Segrè in May 1947, and the α -energies are given in milli-mass-units (m.m.u.).

Plotting again lines of constant E'_α in a $(Z-N)$ -diagram (figure 2), we find that, here too, the E_α -lines perform irregular but more or less parallel wave motions, with the wave-axis approximately parallel to the Weizsäcker line of stable elements. The bulges near $A=19$, 39 and 59 represent minima of the α -binding energy differing by as much as 8 m.m.u. from the normal level of E'_α . The gradients $[\Delta E'_\alpha/\Delta Z]_N$ and $[\Delta E'_\alpha/\Delta N]_Z$ near the minima of E'_α are actually of the same order as those observed with the radioactive elements near $Z=83$, i.e. about 2 mev. per added nucleon.

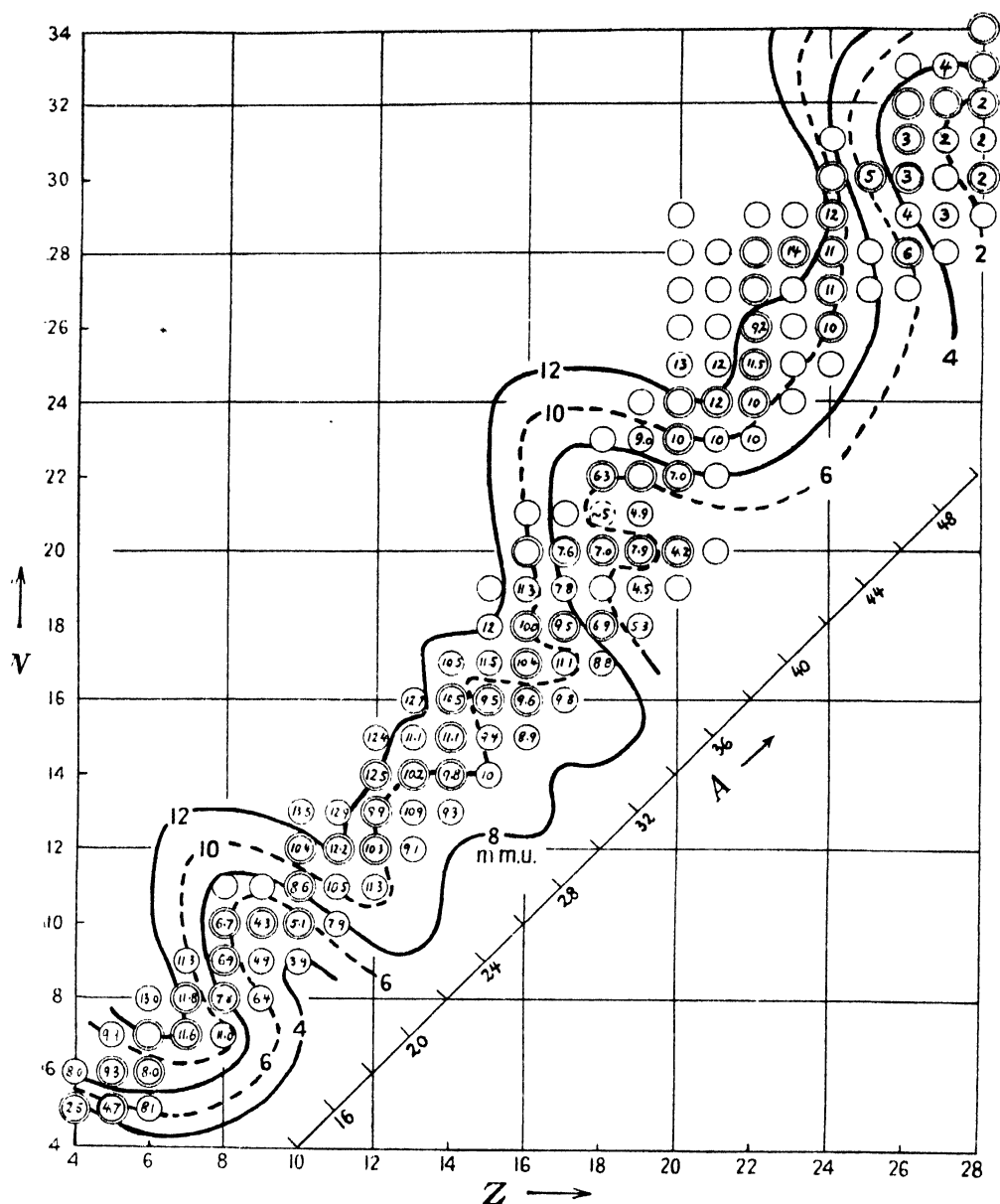


Figure 2. α -binding energies of small nuclei (in m.m.u.).

Double circle: Stable isotopes.

Single circle: Isotopes unstable against β^- , β^+ or K -decay.

(iii) Discussion

There is a considerable temptation to interpret these recurring fluctuations of the α -binding energy, which must be expected also near $A=90$ and 140 (viz. the α -activity of ^{148}Sm), in terms of a "shell" model, though this need not necessarily be of the types considered by Elsasser (1933, 1934) and by Margenau (1934). With a shell structure it appears plausible that the addition of both neutrons and protons to a complete shell has at first an unstabilizing effect.

The α -stability would decrease until the newly added nucleons are numerous enough to interact with each other with a sufficiently large binding energy. From this point onwards the α -stability should increase rapidly to its normal value and remain there—with minor fluctuations—until the shell is filled, after which the process repeats itself.

For small nuclei the electrostatic forces acting on a proton are small, and there is thus not much difference between the addition of a neutron or of a proton. The maxima and minima of E_α should, therefore, be reached at a given number of nucleons independent of whether N or Z or I is kept constant. This is borne out well in figure 2. A similar picture can be applied to the elements of high atomic weight, but here the electrostatic effect of the high nuclear charge, resulting in a tilt of the curve, comes significantly into play; consequently minima of the binding energy (i.e. maxima of the α -energy) are obtained only for lines of constant Z or I , and no longer for constant N .

While this "shell" model must be considered only as tentative, it should be pointed out that it does explain, at least qualitatively, the very puzzling feature that, at the proton numbers $Z=8, 9, 17, 18$, and 26 , the addition of successive neutrons reduces the α -binding energy, and that, at the neutron numbers $N=10, 11$ and others, the addition of successive protons—in spite of their positive charge—increases the binding energy of the last α -particle.

(iv) *The case of argon 39*

The α -binding of ^{39}A , which in figure 2 is given as approximately 5 m.m.u., has been derived in the following way. It follows from the mass difference of $^{35}_{16}\text{S}$ (34.9788) and $^{43}_{20}\text{Ca}$ (43.9723) that $(^{43}_{20}E'_\alpha + ^{39}_{18}E'_\alpha) = 14.3$ m.m.u.

From its position in figure 2 one would expect $^{43}_{20}E'_\alpha$ to be about 9 m.m.u., which leaves for $^{39}_{18}E'_\alpha$ only about 5.3 m.m.u. But even if this value of E'_α for ^{39}A is not accepted, it is apparent from the β -energy of $^{35}_{16}\text{S}$ of only 0.11 m.m.u., and from the fact that ^{39}A must be unstable, that $^{39}_{18}E'_\alpha \leq ^{39}_{19}E'_\alpha$. This fact is of considerable interest, since all theories of the nuclear masses based on a nuclear radius r which grows smoothly with increasing mass ($r \propto A^{1/3}$), require that isobars have an α -binding energy which increases with the isotopic number I (Weizsäcker (1935), Wigner (1937, 1939), or Feenberg (1947)). The case of ^{39}A obviously contradicts this assumption of a smoothly growing radius and requires that a sharp expansion of the nuclear radius takes place in the region of ^{39}A .

To a minor extent this phenomenon occurs also at the other minimum, at $A=17$ and 19 . Here the differences between the E'_α s of isobars for $I=+1$ and $I=-1$ are only about half of the average $\Delta E'_\alpha$ between $A=21$ and 37 , while according to all the theories they should be larger than the $\Delta E'_\alpha$ s of subsequent nuclei ($\Delta E'_\alpha$ approx. $\propto A^{-1/3}$).

In addition to this it is known from the application of the Gamow equation to the data of the decay constants and α -energies of the radioactive elements, that a sharp expansion of the nuclear radius takes place in the region of the elements between $Z=81$ and 83 (Feather 1945), which is exactly where the biggest rise in the α -energies takes place. We therefore conclude that the minima of the α -binding energy shown in figure 2 are also caused by a sudden expansion of the nucleus with increasing mass. While the nuclear radius seems to grow, on the average, with $A^{1/3}$, the possibility must be considered that

it grows only little, if at all, in the interval between two "bulges" of the α -binding energy. This means that the density of the nucleus must be expected to undergo regular fluctuations, minima of the density coinciding with minima of α -stability.

There is some other evidence which supports the conception that the fluctuations in the α -binding energy are not a purely energetic feature but are due to changes of the internal structure.

If the Sargent curve of $\ln E_{\beta+}$ against $\ln \tau$ is plotted for the β^+ -transitions from $I = -1$ to $I = +1$ (figure 3), there is a marked deviation from the otherwise smooth line in the region of $A = 19$. (Such a deviation can also be detected in the function $F\tau$ calculated by Konopinski (1943), which represents essentially the same feature.) It follows that the decay of ^{19}Ne is 2.2 times faster than one would expect from its β^+ -decay energy. There is also a small displacement in the same direction for ^{17}F , ^{41}Sc and most likely also for ^{21}Na .

A similar phenomenon occurs for β^- transitions from $I = +3$ to $I = +1$. In the case of ^{19}O the β^- -decay is about three times slower than one would expect from the extrapolation of the Sargent curve drawn through the points of ^{23}Ne and ^{27}Mg . In either case the nuclei near the minimum of E'_α behave as if the nucleus with the higher isotopic number had a smaller mass than it actually has. The effect reminds us of the similar deviations in the α -decay constants of element 83 in the Geiger-Nuttall plot, which are caused by a comparatively large change of the nuclear radius and density.

§ 2. THE INTERDEPENDENCE OF MASS DEFECT, α - AND β -STABILITY

Feenberg (1947) has shown by semi-empirical considerations that a relationship exists between the mass-defect curve and the β -stability curve of nuclei, and that the irregularities of the one correspond to deviations of the other curve at the same values of the mass-number A . This follows directly—as was shown independently by the author—from the fact that the sum of the α - and β -energies in a cycle must be zero.

It may be seen from figure 4 that

$$4[\Delta E_\beta/\Delta A]_{I \text{ const.}} - 2[\Delta E_\alpha/\Delta I]_{A \text{ const.}} = 0. \quad \dots\dots(1)$$

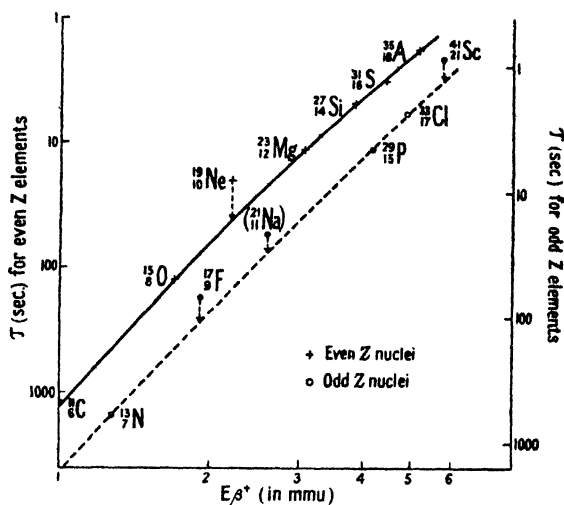


Figure 3. Deviations from Sargent curve near $A = 19$ and 39. (Logarithm of half-life plotted against $\ln E_{\beta+}$).

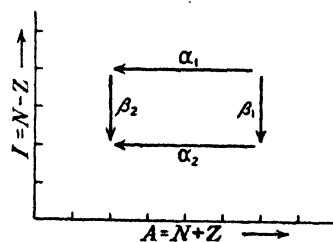


Figure 4. α and β decay cycle.

This can be transformed into

$$\left[\frac{\Delta I}{\Delta A}\right]_{E_\beta=\text{const.}=0} = -\left[\frac{\Delta E_\alpha}{\Delta I}\right]_A / 2 \left[\frac{\Delta E_\beta}{\Delta I}\right]_A = \left[\frac{\Delta E'_\alpha}{\Delta I}\right]_A / 2 \left[\frac{\Delta E_\beta}{\Delta I}\right]_A, \quad \dots\dots (2)$$

where E'_α is the binding energy of the last α -particle in the nucleus, and where $[\Delta I/\Delta A]_{E_\beta=0}$ is the gradient of the line of maximum β -stability, i.e. the bottom of the Gamow valley. $[\Delta E_\beta/\Delta I]_A = [\Delta^2 M/(\Delta I)^2]_A$ is known to be a comparatively smooth and very slowly changing function of A for nuclei of odd mass-number A , which is always positive.

Normally E'_α increases with increasing isotopic number and $I^0 = [I]_{E_\beta=0}$ increases with increasing A .

Equation (2) thus requires that in the region of the well known kinks of the β -stability line, where I^0 decreases with increasing A , the α -binding energy E'_α must also decrease with increasing isotopic number I . Normally E'_α does not change very much for neighbouring nuclei with the same isotopic number, but in order to fulfil this condition it is necessary that E'_α undergo a considerable change within a series of nuclei of the same value of I .

The mass-defect D is related to the α -binding energy by the equation

$$[\Delta D/\Delta A]_{I=\text{const.}} = \frac{1}{4}(D_\alpha - E'_\alpha), \quad \dots\dots (3)$$

where D_α is the mass-defect of ${}^4_2\text{He}$. Equation (3) shows that the mass-defect curve is approximately linear only when E'_α is reasonably constant, and that a marked fluctuation of E'_α must cause a similar deviation of the mass-defect curve from near-linearity.

It thus follows from equations (2) and (3) that in the regions where the β -stability line shows pronounced kinks, as, for example, near the elements ${}_{18}\text{A}$, ${}_{43}\text{Tc}$, ${}_{58}\text{Ce}$ and ${}_{51}\text{II}$, there must be equivalent kinks in the α -stability lines and in the mass-defect curves. Vice versa, in such regions where the α -stability is known to show marked fluctuations, e.g. near ${}_{19}\text{F}$ and ${}_{83}\text{Bi}$, similar deviations must also be expected for the β -stability line and, obviously, for the mass-defect curve.

§ 3. THE ISOBARIC CROSS-SECTION THROUGH THE NUCLEAR MASS SURFACE IN THE REGION OF THE HEAVY ELEMENTS

The preceding study of the α -energies of the radioactive nuclei of large mass has resulted in a diagram (figure 1) which renders possible the interpolation of unknown α -energies with an accuracy of approximately ± 0.1 mev., and a similar, if not better, accuracy for the differences of E_α for isobars (nuclei of constant mass). This makes it possible to calculate also the β -energies (E_β) of unknown nuclei from known β -energies by the relation

$${}^A_Z E_\beta = {}^A_{Z-2} E_\beta + {}^A_Z E_\alpha - {}^A_{Z+1} E_\alpha$$

(E_β stands here for the total energy change in a β -transition equal to ΔM).

The accuracy of the β -energies so obtained depends on the number S of such steps required to obtain the E_β in question, but it should be better than $\pm 0.1 S$ mev. If it were a purely statistical error it would be $0.1 \sqrt{S}$ mev.

This procedure makes it possible to calculate the energies of β -decay (ΔM) of a much larger number of isobars than is obtainable in any other region of the

system of nuclei, not excluding the region of the fission products, and to investigate how far our assumptions concerning nuclear masses are correct, as the β -energy is the gradient of the nuclear mass surface along a line of constant mass-number.

Of particular interest are the isobars of odd mass number A for which the current theories predict a linear variation of ΔM with the isotopic number I .

The calculations have been carried out for the largest and the smallest mass-numbers of the two types of odd A , the mass differences of which can be obtained safely, i.e. 233 and 221 for the series $(4n+1)$ and 239 and 219 for the series $(4n+3)$. These two extremes give a very similar picture and, whilst in the one case the small values of Z and in the other case the large values of Z show the greatest uncertainty, one can be sure that the extrapolation has not caused any significant disturbance.

The results are shown in tables 1 (a) and (b).

The β -energies used for the extrapolations are (in mev.)

for the $(4n+1)$ series :

$${}_{90}^{233} 1.6, \quad {}_{91}^{233} 0.8, \quad {}_{88}^{225} < 0.05, \quad {}_{83}^{213} \sim 1.3, \quad {}_{82}^{209} 0.7;$$

for the $(4n+3)$ series :

$${}_{92}^{239} 1.16, \quad {}_{93}^{239} 0.67, \quad {}_{90}^{231} \sim 0.2, \quad {}_{89}^{227} 0.22, \quad {}_{82}^{211} 1.4, \quad {}_{81}^{207} 1.47.$$

Table 1(a). Type $4n+1$

$A=233$					$A=221$				
Z	I	ΔM	$\Delta^2 M$	Number of steps	Z	I	ΔM	$\Delta^2 M$	Number of steps
90	53	1.6		0	84	53	3.0		3 down
91	51	~ 0.6	1.0	0	85	51	2.1	0.9	3 down
92	49	-1.8	2.4	2 up	86	49	0.6	1.5	1 down
93	47	-2.3	0.5	5 up	87	47	~ 0	0.6	2 up
94	45	-3.6	1.3	6 up	88	45	-1.2	1.2	3 up
95	43				89	43			

Table 1(b). Type $4n+3$

$A=239$					$A=219$				
Z	I	ΔM	$\Delta^2 M$	Number of steps	Z	I	ΔM	$\Delta^2 M$	Number of steps
92	55	1.2		0	82	55	3.9		5 down
93	53	0.1	0.5	0	83	53	3.5	0.4	5 down
94	51	-1.1	1.8	2 up	84	51	1.9	1.6	3 down
95	49	-2.0	0.9	3 up	85	49	1.4	0.5	3 down
96	47	-3.5	1.5	7 up	86	47	0.4	1.0	2 up
97	45	-4.3	0.8	8 up	87	45	-0.5	0.9	3 up
98	43				88	43			

As Δ^2M within the same nuclear type is approximately the same for given values of I , the average of the values of Δ^2M are likely to give a more accurate picture; these are shown in table 2.

Table 2. Average Δ^2M

I	53	51	49	47	45
$A=4n+1$	—	0.95	1.95	0.55	1.25
$A=4n+3$	0.45	1.7	0.7	1.25	0.85

Near $A=229 \pm 10$.

It is apparent that Δ^2M is by no means constant, as required by the various theories (Weizsäcker 1935, Wigner 1937, 1939, Feenberg 1947), but that it varies in such a way that nuclei with an even value of Z always have a larger value of Δ^2M . This means that the mass-defect values of the isobars of odd mass-number plotted against I do not lie on a smooth parabola but on two parabolas, for odd and even values of Z respectively, the latter values being on the average 0.2 mev. smaller than the former. The effect is similar to, but not quite so marked as, that of the nuclei of even mass-numbers, where the greater stability (lower mass) of nuclei with even values of Z is well known.

The trend in the variations of Δ^2M seems to be similar for the $(4n+1)$ and the $(4n+3)$ types of nuclei, and approximately the same values of Δ^2M occur in the case of the $(4n+3)$ types at values of I , which are two units higher than those for the $(4n+1)$ series. It would appear that this effect is another evidence of the formation inside the nucleus of "complexes" consisting of two protons and two neutrons which, in the presence of excess neutron pairs, would result in a somewhat greater stability for nuclei with all protons paired (even Z) than for nuclei with all neutrons paired (odd Z).

ACKNOWLEDGMENTS

The author wishes to thank Professor O. R. Frisch and Mr. J. M. C. Scott, A.E.R.E. Harwell, for many valuable discussions.

REFERENCES

- ELSASSER, J., 1933, *Phys. Radium*, **4**, 549 ; 1934, *Ibid.*, **5**, 389.
 FEATHER, N., 1945, Department of Atomic Energy, Report BR. 640.
 FEENBERG, E., 1947, *Rev. Mod. Phys.*, **19**, 239.
 KONOPINSKI, E. J., 1943, *Rev. Mod. Phys.*, **15**, 209.
 MARGENAU, H., 1934, *Phys. Rev.*, **46**, 613.
 VON WEIZÄCKER, C. F., 1935, *Z. Phys.*, **96**, 431.
 WIGNER, E., 1937, *Phys. Rev.*, **51**, 947 ; *see also* Barkas, W. H., 1939, *Ibid.*, **55**, 691.

The Temperature Effect on Cosmic-Ray Intensity and the Height of Meson Formation

By A. DUPERIER

Turner and Newall Research Fellow of the University of Manchester

Communicated by P. M. S. Blackett ; MS. received 10 March 1948

ABSTRACT. A study has been made of the results obtained in different parts of the world for the effects of ground temperature on cosmic-ray intensity. It is found that the great discrepancies between the values for the temperature coefficient as given by different observers, as well as the seasonal change of this coefficient, far from being a difficulty in Blackett's interpretation may be taken as further evidence of the instability of the meson. The 12-monthly variation in cosmic-ray intensity as obtained by Forbush can be also accounted for as an effect of meson decay if one assumes that the bulk of the penetrating component originates at 7.5 cm. Hg pressure-level. This assumption has been put to the test and the results obtained appear to confirm that the layer of meson formation lies at, or little above, the 7.5 cm. pressure-level, at the height of the maximum of the Pfitzner curve for the intensity of the total radiation.

§ 1. INTRODUCTION

IN the last fifteen years, it has been reported by a number of investigators that the intensity of cosmic rays at constant atmospheric pressure shows from day to day and from month to month a variation inverse to that of the air temperature near the ground. It is well known that Blackett (1938) attempted to relate this temperature effect to the instability of the meson, and that in this way he explained the decrease of cosmic-ray intensity in warm weather as due to the greater height of the pressure level at which mesons originate. As Blackett pointed out, the coefficient of the temperature effect at the equator should be lower than at moderate latitudes since the incident primary rays are more energetic there.

However, when the experimental results obtained in different parts of the world are closely examined, several difficulties in this interpretation appear to arise. We have first the fact that the values of the temperature coefficient for stations at moderate latitudes, as obtained by different observers, show discrepancies among themselves which are much greater than would have been expected from the accuracy of the measurements. These discrepancies are particularly noticeable if we compare the values obtained by following one method of computation with those obtained by another. Some observers have correlated the monthly or seasonal means of cosmic-ray intensity with monthly or seasonal means of ground temperatures. Others have computed the coefficient for each month separately by using the daily means and then averaging the twelve coefficients. The former method appears to lead to values which are about twice as great as those obtained by the latter.

In the second place, when computed for each month separately, the temperature coefficient shows a most unexpected seasonal variation.

Forbush (1938) has shown that the annual variation in cosmic-ray intensity consists of two components—one of a world-wide character related to geomagnetic activity, and the other a 12-monthly wave inverse to that of the

temperature. There might be added as a further difficulty in the interpretation by meson decay the fact that the comparison of this 12-monthly wave with the annual change in ground temperature leads to a temperature coefficient which varies from one station to another.

In a former paper (Duperier 1941) the author showed, though in a manner which could not be considered as conclusive, that some at least of these difficulties would disappear if the mean temperature of the atmosphere, instead of the temperature near the ground, had been used for the correlation with cosmic-ray intensity.

The more complete information concerning the upper air, which has been gathered in the last few years in England by means of four daily radio-sondes from the same station, has made possible a further study of the extent to which the temperature effect may be related to the instability of the meson.

§ 2. THE TEMPERATURE EFFECT

In a previous publication (Duperier 1945) it was reported that the intensity of cosmic rays, as measured by a counter arrangement registering triple coincidences, appears to be closely related to the height of the 7.5 cm. Hg pressure-level. It was found that the correlation of cosmic-ray intensity at constant atmospheric pressure with the height of lower pressure-levels is rather low, though gradually increasing with height. When the 7.5 cm. pressure-level is chosen—the highest for which sufficient meteorological data was then available—this correlation reaches the values -0.67 . By admitting the principle of the instability of the meson and assuming that the bulk of the penetrating component originates at this pressure-level, the value 18.6 km. was obtained for L , the mean range of mesons.

Now, if δh represents the change in height corresponding to the variation δt of ground temperature, the coefficient α of the ground temperature effect will be given by the expression

$$\alpha = -(1/L)(\delta h/\delta t). \quad \dots\dots (1)$$

As the heating of the atmosphere by conduction and convection, and so the expansion δh corresponding to a certain variation δt of the ground temperature, is the greater the longer the period during which this variation occurs, it is to be expected that the value of the ground temperature coefficient obtained by using variations from month to month will be greater than that obtained by the use of variations from day to day.

In the last few years it has been the practice in England to make radio-sonde observations every day at 00, 06, 12, 18 hours G.M.T. For the present study, the upper-air data obtained at Larkhill (about 100 km. to the south-west of London) during the period November 1943–October 1945 have been used, and the heights of the 7.5 cm. pressure-level determined.

The mean heights at 00, 06, 12, 18 hours for each season have been plotted in figure 1, together with the mean temperature near the ground.

It can be seen that, contrary to general belief, the mean height for each season changes quite appreciably with the time of day. As the number of days used for the computation of the means (more than 2000 in all) is sufficiently high practically to eliminate the effect of change in ground pressure, the variation in height of the 7.5 cm. pressure-level has to be ascribed to the thermal structure of the air below it.

To see, however, how closely the diurnal variation in height is related to the diurnal variation of ground temperature alone, departures from daily mean height for each hour have been correlated with similar departures of ground temperature. By doing this, the seasonal effect is obviously eliminated, and we find a value of 0.92 ± 0.02 for the correlation, and 18 ± 1 metres per $^{\circ}\text{C}$. for b_1 , the regression coefficient.

If we substitute b_1 for $\delta h/\delta t$ and 18,600 m. for L in (1), we have for the temperature coefficient $\alpha_1 = -0.10\%$ of mean intensity per $^{\circ}\text{C}$. This value can be compared with those given by Hess *et al.* (1940) and by Hogg (1947), since these

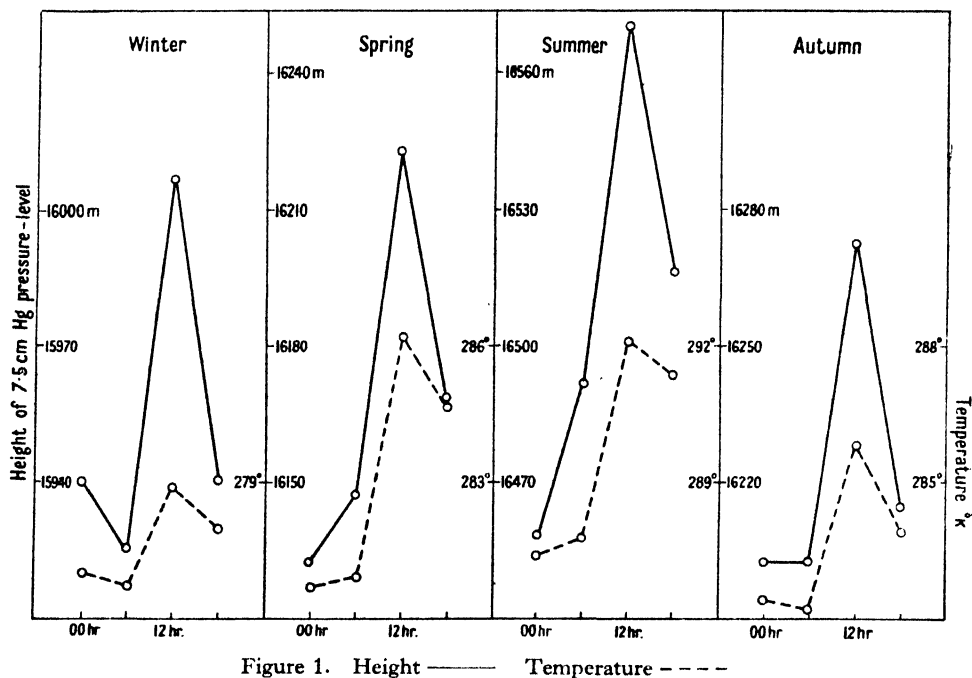


Figure 1. Height ——— Temperature - - - -

observers have followed a method of computation by means of which the seasonal effect is also eliminated. From series of observations of cosmic-ray ionization covering periods of several years, they obtained α_1 for each month separately by using daily means, and then averaging the twelve coefficients. As table 1 shows, the three values are the same.

Table 1

Locality	α_1 (%)	Observer
The Hafelekar (Austria)	-0.09 and -0.107	Hess <i>et al.</i> (1940)
Canberra (Australia)	-0.11	Hogg (1947)
London	-0.10	Duperier
Mean	-0.10	

Hess *et al.* and Hogg have also reported the existence of a marked seasonal change of the temperature effect. Hess has found that α_1 is nearly twice as great in winter as in summer. From Hogg's results it can be seen that the ratio of the two coefficients is 1.5.

That the change in height of the 7.5 cm. pressure-level is more pronounced in winter than in summer, compared with that in ground temperature, is already apparent in figure 1. From the regression coefficients for the winter and summer groups separately, the value 1.6 is obtained for the ratio α_{1W}/α_{1S} , which is quite consistent with those of Hess and Hogg.

If we correlate monthly averages of the means of the four heights and temperatures for each day (plotted in figure 2) instead of correlating daily departures, then we obtain for the correlation and regression coefficients the values 0.98 ± 0.01 and 45 ± 2 metres per $^{\circ}\text{C}$. respectively. By applying the equation (1), we have now for the temperature coefficient $\alpha_2 = -45/18,600 = -0.24\% \text{ per } ^{\circ}\text{C}$. which is about double the value of α_1 obtained before.

The temperature effect obtained from monthly means should therefore be about twice as great as the value from daily means.

Table 2 gives the coefficients which have been obtained in different parts of the world by using monthly or seasonal means of cosmic-ray ionization and ground temperature.

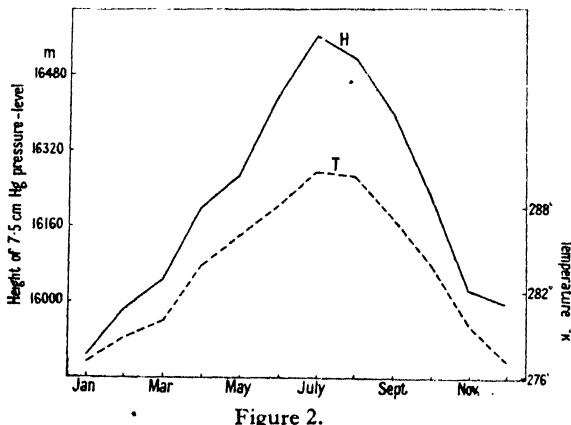


Figure 2.

Table 2

Locality	α_2 (%)	Observer
Voyage Vancouver (Canada)	{	Compton and Turner (1937)
to Sydney (Australia)		Gill (1939)
Cape Town		Schonland <i>et al.</i> (1937)
Amsterdam		Clay and Bruins (1939)
Mean	-0.18	

It can be seen that all the values are greater than those given in table 1 and that the mean is about twice as great, in agreement with the results from the radio-sonde observations at Larkhill. The relatively small value for Cape Town, at 32.7°S . geomagnetic latitude, may be partly due to the primary radiation being more energetic there.

On the hypothesis of the instability of the meson, the use of daily means results in a smaller value for α on account of the lag in the warming of the atmosphere relatively to the warming of the ground, as previously stated. As the height of any pressure-level actually depends on the spatial temperature defined by $\int_1^z t(z) dz/z$ of the air below it, it is clear that had the cosmic-ray intensity been compared with this temperature, the result would have been the same independent of the length of the period over which the data are averaged.

In conclusion, the discrepancies between the temperature coefficients as obtained by different observers are due at least for the most part to the methods of computation having been different. Far from being a difficulty in the meson

decay interpretation, these discrepancies, together with the seasonal change of the ground temperature effect, may be taken as further evidence of the instability of the meson.

§ 3. THE 12-MONTHLY VARIATION

Figure 2 shows the annual change in height of the 7.5 cm. pressure-level. The harmonic analysis leads to a 12-monthly variation with an amplitude of 307 metres and the maximum in mid-summer. If again we assume that the bulk of mesons originate at this pressure-level we should have at the latitude of London a 12-monthly variation in cosmic-ray intensity with an amplitude of $307/18,600 = 1.65\%$ of mean intensity, and the maximum in mid-winter.

Forbush has shown from the analysis of the observations made at five widely separated stations that the annual variation of cosmic-ray intensity consists of two components—one of a world-wide character corresponding with geomagnetic activity, and the other a 12-monthly wave inverse to that of the temperature near the ground. Table 3 gives the amplitudes and times of maxima of this 12-monthly wave as obtained by Forbush for The Hafelekar, Cheltenham and Christchurch, the latitudes of which are comparable to that of London.

Table 3

Station	Latitude	Geomag. lat.	12-monthly variation	
			Ampl. (%)	Maximum
Cheltenham (U.S.A.)	38.7° N.	50.1° N.	1.6	January 19
Christchurch (N.Z.)	43.5° S.	48.0° S.	0.8	July 28
Hafelekar (Austria)	47.3° N.	48.4° N.	1.9	January 15
		Mean	1.43	
London (predicted)	51.5° N.	54.0° N.	1.65	January 18

It can be seen that the predicted amplitude and phase for London is in fairly good agreement with the mean value of the amplitudes and phases at the other three stations.*

As the upper air observations show that for stations at similar latitudes the range of the annual variation in atmospheric temperature, in contrast with the range of the variation of the ground temperature, is but little affected by the climate, it may be taken that the variation in height of the 7.5 cm. pressure-level in London is comparable with the mean of these variations at the other stations. If this is so, the agreement shown by the table may be taken as indicating strongly that the 12-monthly variation in cosmic-ray intensity is solely the effect of meson decay.

This conclusion seems to rule out the possibility of explaining the observed 12-monthly variation as due to the heliomagnetic field, as suggested by Vallarta and Godart (1939). If this field does not give rise to such a variation, the reason is perhaps that the numbers of positive and negative particles in the primary radiation are equal or, alternatively, it may be that the properties of the permanent magnetic field of the sun are such that this field does not change the intensity of the incoming radiation in the course of the year.

* Yearly variation not determinable by counter-apparatus.

§4. THE HEIGHT OF MESON FORMATION

We have seen in the preceding sections that the temperature effect as well as the 12-monthly variation of cosmic-ray intensity, as obtained in different parts of the world, are well explained by assuming that the bulk of mesons originate at the 7.5 cm. Hg pressure-level.

In addition to this we have the fact that the value of 18.6 km. for the mean range, obtained by the same hypothesis, leads to a rest life which agrees with the one generally accepted (Nereson and Rossi 1943). If we take for the mass $180 m_e$, and for the relevant average momentum of the mesons along their path down to sea-level the value 2800 mev/c. given by Rossi (1939) (derived from Blackett's spectrum and assuming a momentum loss of 2×10^6 ev/c. per gm/cm²), then from $L = p\tau_0/\mu$ we have $\tau_0 = 2.04 \times 10^{-6}$ sec.

We might therefore conclude from these results that the bulk of mesons is generated at the 7.5 cm. pressure-level.

To see, however, what the results would be if a higher pressure-level had been assumed for the layer of meson formation, all the preceding computations have been repeated taking the 3 cm. pressure-level, the highest for which sufficient meteorological data are now available. Table 4 compares the results so obtained with the previous results.

Table 4

	α_1 (%)	α_2 (%)	12-monthly var. ampl. (%)
3.0 cm. Hg pressure-level	- 0.15	- 0.43	2.40
7.5 cm. Hg pressure-level	- 0.10	- 0.24	1.65
Observed values	- 0.10	- 0.18	1.43

We can see that the hypothesis of meson formation at the 3 cm. pressure-level leads to values which are roughly 50% higher than those for 7.5 cm. Hg. To reconcile them with the experimental results, we should clearly have to take for the mean range of mesons a value of 28 km., a 50% increase, and this value, together with the fact that in this case we should have to take for the average momentum of the mesons along their path 2100 mev/c. (Jánosy 1948), would lead to a rest life of 4.0×10^{-6} sec., a 100% increase over the accepted value.

The conclusion appears therefore to be that the bulk of mesons is generated at, or little above, the 7.5 cm. pressure-level, at the mean height of 16 km.

If the average layer of meson formation lies at that height, it follows that the intensity of the meson component should be decreasing at 3 cm. Hg (22 km.). On the other hand, the experimental result of Schein, Jesse and Wollan (1941) shows that the hard component increases gradually up to the height of 2 cm. Hg reached in their experiment. The two results, however, are not necessarily mutually exclusive if, as is generally supposed, the mesons are formed by primary protons the penetrating power of which is not too different from that of the hard component. The intensity of the primary protons should increase gradually with height.

In connection with the origin of the soft component, it may be significant that the layer of meson formation appears to be located at the height of the maximum of the Pfotzer curve for the total intensity (1936).

ACKNOWLEDGMENTS

The author wishes to express his indebtedness to the Director of the Meteorological Office, Air Ministry, for kindly placing the upper-air data at his disposal.

He also wishes to express gratitude to Sir George Thomson for providing facilities for this work to be carried out in the Physics Department, Imperial College of Science and Technology, and to Professor P. M. S. Blackett for his very kind interest and helpful suggestions at all times.

REFERENCES

- BLACKETT, P. M. S., 1938, *Phys. Rev.*, **54**, 973.
 CLAY, J., and BRUINS, E. M., 1939, *Rev. Mod. Phys.*, **11**, 158.
 COMPTON, A. H., and TURNER, R. N., 1937, *Phys. Rev.*, **52**, 799.
 DUPELIER, A., 1941, *Proc. Roy. Soc. A*, **177**, 204 ; 1945, *Proc. Phys. Soc.*, **57**, 464.
 FORBUSH, S. E., 1938, *Phys. Rev.*, **54**, 975.
 GILL, P. S., 1939, *Phys. Rev.*, **55**, 1154.
 HESS, V. F., 1940, *Phys. Rev.*, **57**, 781.
 HOGG, A. R., 1947, *Proc. Roy. Soc. A*, **192**, 128.
 JÁNOSSY, L., 1948, *Cosmic Rays* (Oxford : University Press).
 NERESON, N., and ROSSI, B., 1943, *Phys. Rev.*, **64**, 199.
 PFOTZER, G., 1936, *Z. Phys.*, **102**, 23.
 ROSSI, B., 1939, *Rev. Mod. Phys.*, **11**, 296.
 SCHEIN, M., JESSE, W. P., and WOLLAN, E. O., 1941, *Phys. Rev.*, **59**, 615.
 SCHONLAND, B. F. J., DELATIZKY, B., and GASKELL, J., 1937, *Terr. Mag. and Atmos. Elect.*, **42**, 137.
 VALLARTA, M. S., and GODART, O., 1939, *Rev. Mod. Phys.*, **11**, 180.

The Attenuation of Ultra-High-Frequency Electromagnetic Radiation by Rocks

By R. I. B. COOPER,

Department of Geodesy and Geophysics, University of Cambridge

Communicated by B. C. Browne ; MS. received 21 November 1947, and in amended form 25 February 1948

ABSTRACT. Radiation of frequency 200 Mc/s. was transmitted through the covering rocks of certain railway tunnels from a radar transmitter feeding a special aerial. The object was to measure the attenuation of rocks *in situ* to see if there might be practical applications to geophysical prospecting. The attenuation in dry sandstone was found to be 3 to 4 db. per foot, which was much greater than had been hoped. It thus appears that the penetration is insufficient to be of practical use.

§1. INTRODUCTION

THE use of electromagnetic radiation for geophysical prospecting, as opposed to the use of induction methods, does not appear to have been extensively discussed. A great deal of literature exists, however, on the effect of the surface conductivity of the ground on the propagation of surface waves. Recently in America attention has been focused on the possible use of surface wave measurements for detecting dykes and faults or other concealed geological discontinuities which have a considered effect on the signal strength measured near them. In Germany Grosskopf and Vogt (1940, 1942, 1943) have been principally responsible for the interpretation of effective surface conductivities in terms of a structure consisting of layers of different conductivity. In Russia, Mandelstam and

Papalex (1938) have developed a technique which was used by Al'pert (1945) to investigate the refraction effects near a horizontal conductivity boundary.

By contrast, little attention has been drawn to the possibility of penetrating the earth by suitably directed radiation. It has usually been regarded as impossible because of the conductivity of the rocks; nevertheless, consultation of available figures suggested that some rocks might be found with sufficiently low conductivity. Jakosky (1937), dealing quantitatively with the effect of moisture content on the conductivities of different rocks, found that the conductivity depended principally on the amount of moisture present, and very little on the material of the rock. A sandstone with 2–3% water was found to have a conductivity of about 10^{-14} E.M.U.* Figures for effective surface conductivity given by Grosskopf and Vogt ranged from $9 \cdot 10^{-15}$ to $4 \cdot 10^{-13}$, and they suggested that the rocks below the upper layer might have smaller conductivity. Fritsch (1943) performed some experiments on radio communication in mines at frequencies between 4 and 30 Mc/s. and concluded that the ranges achieved accorded with a smaller attenuation than was commonly expected.

The simplest way of using electromagnetic radiation for prospecting would be to measure the time required for a pulse to be transmitted to a buried object and reflected. For such a "radar" technique the frequencies used must be high enough to enable sufficiently short pulses to be generated. Setting the pulse length arbitrarily at $\frac{1}{5} \mu\text{sec.}$ it is seen that frequencies much below 200 Mc/s. are useless. The U.H.F. conductivity loss is nearly independent of frequency, but the dielectric losses, notably of water, increase rapidly with increasing frequency: at 200 Mc/s. these losses are still negligible in comparison with the conductivity loss, and the losses of other constituents in the main body of rock—mica, quartz, calcite, feldspar, etc.—are still small compared with the loss in the water present.

A calculation given later shows that, for a conductivity of 10^{-14} and a frequency of 200 Mc/s. reflections might be obtained from within about 50 m. Losses due to scattering, imperfect reflection, aerial mismatches, etc., would affect this range, but since it is determined mainly by an attenuation it would not be reduced appreciably provided these losses did not exceed a few db. On the other hand, the range would be almost exactly inversely proportional to the conductivity. Nevertheless, in view of the lack of data and the importance of the possibilities, an experiment at a selected dry site was considered practicable.

To determine the attenuation without using reflections, a radar transmitter was mounted above a railway tunnel and provided with a special aerial for directing radiation into the ground after the removal of the surface soil. An exploring receiver was used to measure the field in the tunnel, the output being displayed on a cathode-ray tube with a high-speed trace. The tunnels investigated were: Colwall, Worcs., keuper marl; Alderton, Glos., great oolite; and Pontypridd, Glam., coal measure sandstone.

§ 2. EXPERIMENTAL METHOD

Essentially the method consisted in transmitting a known power downwards into the ground above the tunnel and measuring the field strength with a receiver in the tunnel after passage of the radiation through a given thickness of rock.

* E.M.U. are used throughout the paper unless otherwise stated.

The field strength E at the receiver is given approximately by

$$E(\text{v/cm.}) = (7.0\sqrt{(PGf_T f_R)/r}) \times e^{-kt}, \quad \dots\dots(1)$$

P =power of transmitter (watts); f_T =fraction of power entering ground; f_R =fraction of power emerging from rock to reach receiver; G =aerial gain *; t =thickness of rock (cm.); r =distance from transmitter to receiver (cm.); k is defined as the attenuation constant and $1/k$ is seen to be the distance in which the field strength is reduced to $1/e$ of the value it would have in the absence of the rock ($k=0$ in *vacuo*).

It can be shown that $2k^2 = (4\pi^2/\lambda^2)[\sqrt{(\epsilon^2 + 4\sigma^2/f^2)} - \epsilon]$, where λ =wavelength; ϵ =dielectric constant; σ =conductivity in E.S.U.; f =frequency.

At 200 Mc/s. we find, for reasonable values of ϵ and σ , that $\epsilon \gg 2\sigma/f$ and

$$1/k \simeq c\sqrt{\epsilon}/2\pi\sigma; \quad \dots\dots(2)$$

with $\epsilon=9$ and $\sigma=10^{-14}$ E.M.U.

$$1/k \simeq 16 \text{ metres.}$$

The exponential factor in (1) is thus seen to be the governing term in determining the range of penetration to be expected and that, therefore, when solving for r , orders of magnitude only need be inserted for quantities in this expression other than k . f_T may be assumed 0.5 since we can control the method of direction into the ground, but f_R may be only of order 0.1 since for practical reasons it is impossible to provide a special receiving aerial inside a railway tunnel. Putting $E=10^{-6}$, $PG=10^5$ and $r=t$, we then find $r \sim 150$ metres. This distance will, however, be almost exactly inversely proportional to σ . In the event of the attenuation being low enough for reflections to be obtained, the expected maximum depth of detection would be rather less than half this figure. It is not expected that ϵ will show very wide variations and the expected range depends in any case only on its square root.

In order to determine k , the variation of E with t should have been studied. This would have involved serious practical difficulties. Either the rock would have had to be removed from the site to reduce the thickness between transmitter and receiver or else new sites would have had to be occupied with a thinner cover to the tunnel. The former alternative would have been prohibitively laborious and the latter procedure would have been invalid if the rocks were inhomogeneous in a horizontal direction; the terrain might also have proved to be unsuitable. Therefore it was resolved to measure each of the quantities in (1) separately at any site which proved suitable. It must be borne in mind that the main aim of the experiments was to discover *rapidly* whether there was any chance of the proposed technique being successful: had it so proved, more accurate measurements would have been undertaken.

Measurement of E , t and r was simple, and a theoretical figure was used for G ; the measurement of P , f_T and f_R presented difficulties. It is clear that in endeavouring to direct power from a transmitter in air into the ground there will be loss due to reflection at the interface. For the purpose of these experiments the magnitude of this loss was immaterial so long as it could be measured. Therefore a special aerial was constructed (see § 3) from which power could not escape except either by entering the ground or returning down the lead to the transmitter. The power flowing both to and from the aerial was measured and the difference

* This is the ratio of the power per unit solid angle sent in the direction of the receiver to that which would be sent if the power were uniformly radiated in all directions.

therefore represented the amount entering the ground. In this way the values of P and f_T were found.

f_R was simply the transmission coefficient from the rock into the air in the tunnel since no special aerial was used. It was assumed to be the same as that at the surface where the transmitter was situated, and was deduced from measurements made on the transmitting aerial. An overall test was also made in free space by transmitting across a stretch of water and comparing the observed value of E with that calculated from P , G and r . This test gave satisfactory agreement.

§3. DETAILS OF THE EXPERIMENTAL EQUIPMENT

A block-schematic diagram of the equipment is given in figure 1.

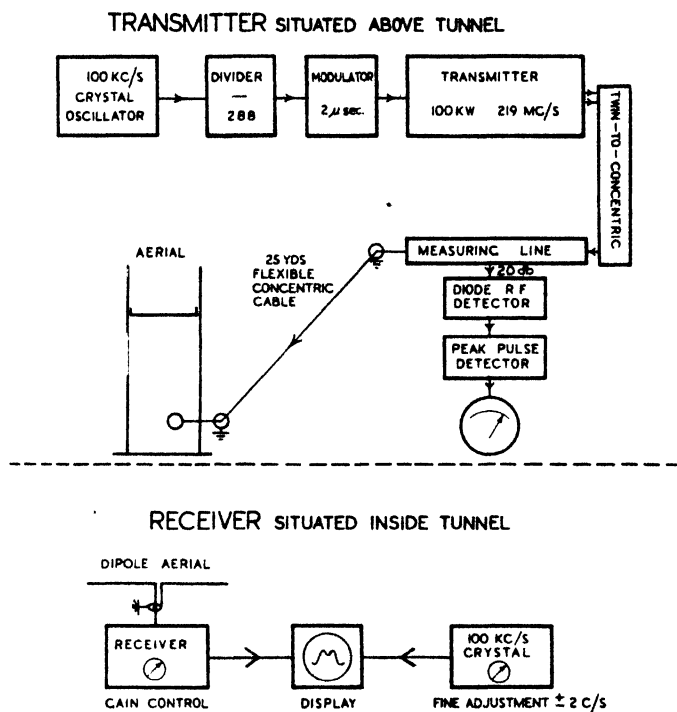


Figure 1. Block schematic diagram of apparatus.

(i) Transmitter

Type T3154B; frequency 219 Mc/s.; nominal peak power 100 kw.; pulse length $2\mu\text{sec.}$; recurrence 3–400 c/s.

It was necessary to measure the peak power being actually transmitted into the ground and to monitor this continuously in the field. Since no wattmeter operating at this frequency was available, a concentric-line directional coupling unit was developed. This type of measuring line has only been described in Service reports; a small amount of the power passing along the line in one direction only can be coupled into a suitable detector. The transmitter output was converted from balanced twin feeder to concentric line by a suitable transformer before connection to the measuring line: from the latter, 25 yds. of flexible cable (Uniradio 34) led to the aerial. The measuring line was double-ended; it could

thus be reversed, so that the power flowing both to and from the aerial could be read. The peak power at the transmitter was usually about 75 kw.

(ii) Transmitter aerial

This consisted of an open-mouthed waveguide of section 45×15 in. about 3 ft. long and provided with a movable reflecting piston and adjustable probe, connected by concentric cable (Uniradio 34) to the end of the measuring line (figure 2). The cable attenuation was 1.8 db. The best dimensions of the probe were first found by pointing the mouth upwards into free space and altering the probe length and piston-probe spacing to achieve the best match. As the dimensions affected only the matching of the cable into the impedance of the guide, they were not subsequently changed. The mouth was then rested on the ground and the positions of both probe and reflector adjusted to give the best match. Measurements showed that usually 75% of the power available at the probe entered the ground, and this was considered adequate.

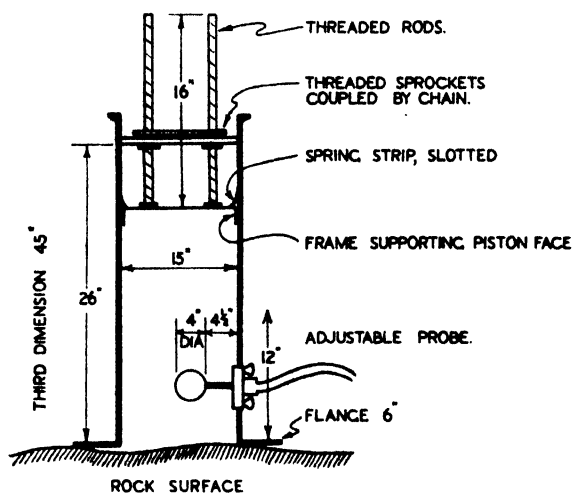


Figure 2. Waveguide aerial.

(iii) Receiver

Type R 3170; frequency 219 Mc/s.; bandwidth 1.4 Mc/s.; noise level 0.5×10^{-12} w.

It was necessary to provide a rapid means of monitoring the receiver noise-level in the field. For this purpose a noise-source was constructed using a CV 172 diode and run from accumulators (B.T.H. report). Measurements showed that its output near 219 Mc/s. was 3 db. less than the theoretical value.

The receiver gain was calibrated by well known methods (Cooper and Freeman 1943, Varley *et al.* 1943).

(iv) Receiver aerial and display

The receiver aerial was a simple dipole mounted on a bakelite rod and fed from a few yards of flexible cable so that it could be used to explore the field. Its matching into the receiver was determined with the aid of the measuring line. It was found that at most 10% of the power was reflected.

In order to display the received pulses without introducing a connection between transmitter and receiver down which radiation might leak, and at the same time to have a time-base of sufficiently high speed to enable small time delays to be detected, it was necessary to control the recurrence of both transmitter and receiver from crystals. The transmitter recurrence was derived from a 100-kc/s. crystal by repeated division with multivibrators, and the receiver time-base

consisted of a triangular waveform with a forward stroke of $7\mu\text{sec}$. directly generated from another 100-kc/s. crystal. A small trimmer enabled the display to be synchronized with the transmitter to within $\frac{1}{10}$ c/s. In this way the time-distribution of events could be studied sufficiently exactly to distinguish the direct signal from that leaking round the tunnel mouth.

§ 4. RESULTS

The selection of the tunnels was a matter of some difficulty. The main consideration was that they should be dry, but they also had to have convenient access by road, both for the transmitter in its vehicle above the tunnel and for the receiver and its generator, which had to be run in on a trolley from the nearest station. Traffic density was also a factor in the selection.

The dryness of the tunnels could only be judged by inspection of the walls and the ground above, but this did not necessarily mean that water was not present in the intervening rock. The experiments were conducted in August 1947, which was the driest period experienced for several decades. However, neither in Colwall nor Alderton tunnels were any indications obtained of signals arriving other than by leakage round the mouth: it can only be supposed that the intervening rocks were wetter than expected. The cover in both cases was about 40 ft. thick.

In Colwall tunnel sites were tried at both western and eastern ends; the western end was rather wet (a spring was later discovered near the tunnel mouth), whilst at the drier eastern end the transmitter could be mounted only on the road immediately above the line of the tunnel and about 35 yds. back from the mouth, where considerable radiation leakage was experienced. This made the field in the tunnel very complicated. The traffic-density in this single-track tunnel made thorough exploration of the field difficult, but no obvious indications of direct transmission were noticed. As the keuper marl at this end is at the foot of the steep Malvern Hills it would not be surprising if it were damper than superficially appeared.

In Alderton tunnel the cover is constant and about 40 ft. thick, and consists of very inhomogeneous intercalations of limestone and clay. Although this appeared very dry, it is quite possible that much scattering and multiple reflection could occur. Two sites were tried immediately above the line of the tunnel, and about 100 yds. from the eastern end, to see if accidental imperfections were affecting the penetration, but no success was obtained at either.

It is clear that the efficiency of transfer of the radiation into the air in the tunnel from the rocks is low. Not only is the transmission coefficient small, but the curved tunnel roof acts as a lens tending to spread the beam and reduce the field strength. From measurements on the transmitting aerial a transmission coefficient of 0.2 was obtained, and from this the refractive index of the ground could be estimated: it was thus possible to calculate the "lens effect" from the dimensions of the tunnel. In the case of Alderton and Pontypridd tunnels the total calculated loss was about 25 db.

At Pontypridd the site was considerably better: the cover increased gradually so that at the site 100 yds. inside the northern mouth it was only 33 ft. thick; it consisted of massive homogeneous dry sandstone and the tunnel was almost disused, so that it was possible to run the receiver up and down the line on a trolley and observe and plot the signal strength without waiting for the passage of trains and clearance of smoke.

In figure 3 is shown the variation with distance into the tunnel of the signal strength, assessed as the height of the largest pulse visible. It is believed that the full curve shows the decay of the signals leaking into the mouth, and the dotted curve that of signals received directly through the roof. It will be seen that the observations show a greater scatter in the region below the transmitter and

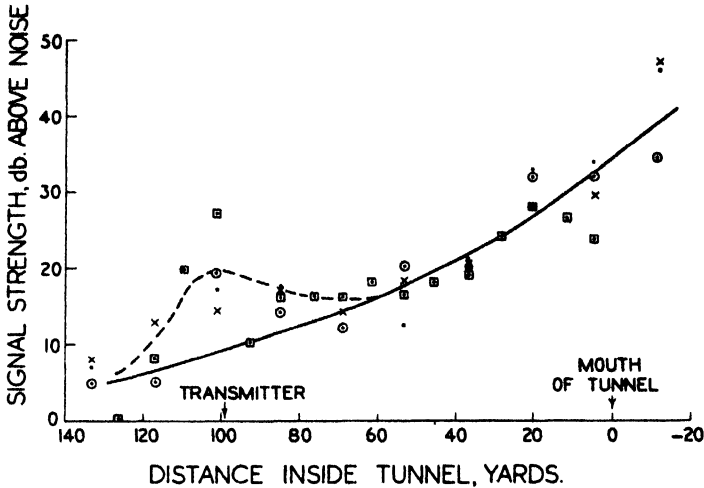


Figure 3. Variation of signal strength in Pontypridd tunnel.

- 1st run, Sept. 9. Polarized $\parallel$ rails.
- 2nd run „ „ „ $\perp$ „
- × 3rd run Sept. 9. Polarized $\parallel$ rails.
- 4th run „ 10. „ $\parallel$ „

near the mouth: this may be because here there are strong interference fields, and small movements of the aerial, such as might easily occur in between runs, can cause large changes in signal strength. Nevertheless it is considered that the small maximum below the transmitter is significant (the distance of the transmitter from the mouth was not known until after the measurements had been completed). The aerial was mounted about 3 ft. above the offside rail: in the non-interfering zone the strength did not depend in a regular manner on position or orientation provided the aerial was not too close to floor or walls. Runs were made with the aerial polarized horizontally both perpendicular and parallel to the rails, but no outstanding difference was observable.

The trace was also observed whilst the trolley was run back and forth under the transmitter site: a particular echo on the trace was observed to grow and decay during this test, which was very convincing. The appearance of the trace is sketched in figure 4. A small pulse is seen before the direct signal, which should in this case be the earliest arrival, but since the time-base runs continuously this may in fact be an echo arriving on the next trace, $10\mu\text{sec.}$ later. This would be possible since many large reflecting objects were situated at the required range from the tunnel mouth. Ignoring this pulse, therefore, the next pulse seen in figure 4 is delayed by approximately $1\mu\text{sec.}$ on the direct pulse,

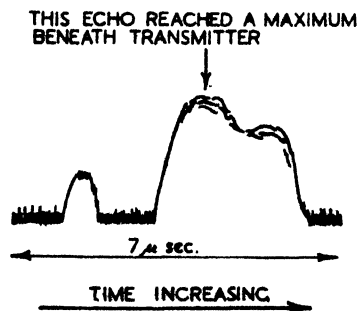


Figure 4. Appearance of trace with signals.

or about 350 yds.—a distance of the right order for a path round the tunnel mouth.

The signals were strongly modulated by the approach of trains.

§5. CALCULATION OF ATTENUATION

The transmitter power is $7.5 \cdot 10^4$ w. and aerial gain ~ 5 ; the power flux at 1 yd. distance is therefore $7.5 \cdot 10^4 (5/4\pi) (91.44)^2 = 3.56$ w/cm². This is equivalent to a field strength of 69.4 v/cm.

The observed signal in the tunnel at 16 yds. range was 10 db. above noise. The noise-level of the receiver at the time of the experiment was $4 \cdot 10^{-13}$ w.

A field of strength f v/cm. produces power $\lambda^2 f^2 / 320\pi^2$ w. in a receiver matched to a simple dipole. Hence in this case, since $\lambda = 135$ cm., we find $f = 0.838 \cdot 10^{-6}$ v/cm. The field expected in free space at 16 yds. range is $69.4/16 = 4.34$ v/cm. Hence the attenuation to be accounted for is $4.34/0.838 \cdot 10^{-6}$ or 134.3 db. Allowing 3 db. loss at the transmitter and 25 db. for the effects where the wave enters the tunnel, we find the attenuation of Pennant sandstone to be 3.2 db/ft. This is about 20 times greater than that expected assuming $\sigma = 10^{-14}$ E.M.U., so that the conductivity must have been about 2×10^{-13} E.M.U. (resistivity 5000 Ω /cm.). According to Jakosky this could have been accounted for by the presence of 15% of water, assuming Pennant sandstone to be similar to his reservoir sandstone. It seems therefore that even after prolonged surface dryness the rocks near the surface may be quite damp.

§6. CONCLUSION

The penetration of ultra-high-frequency electromagnetic radiation into rocks is insufficient to afford a practical method of geophysical prospecting. This is probably because of the great power of rocks to occlude and retain moisture even after prolonged drought. In cases where there was appreciable scattering due to inhomogeneous ground there would be even less chance of success.

Increase of power and gain would not increase the range greatly since the attenuation is exponential. Sufficient penetration could not be obtained by reduction of frequency since pulse widths would then be too great for the observation of reflections.

ACKNOWLEDGMENTS

The success of the experiments was largely due to the full co-operation of the G.W.R., and the author wishes to express here his sincere gratitude to them. The radar set was lent by kind permission of the Radar Research and Development Establishment, Ministry of Supply, Malvern, who gave laboratory facilities for the calibrations, and to whom also the author is very much indebted.

REFERENCES

- AL'PERT, J. L., and GOROZHANKIN, H., 1945, *J. Phys.*, U.S.S.R., 9, 115.
 BIRCH, F., 1942, *Handbook of Physical Constants*, Geol. Soc. Amer. Spec. Paper No. 36.
 B.T.H., Special Report No. L293S.
 COOPER, R. I. B., and FREEMAN, J. Y., R.R.D.E., Ministry of Supply, Research Report 251.
 FRITSCH, V., 1943, *Hochfrequenztech. u. Elektroakust.*, 62, 50.
 GROSSKOPF, J., 1942, *Hochfrequenztech. u. Elektroakust.*, 60, 136.
 GROSSKOPF, J., and VOGT, K., 1940, *Telegr.- u. Fernspr. Tech.*, 29, 164; 1943, *Hochfrequenztech. u. Elektroakust.*, 62, 14.
 JAKOSKY, J. J., 1937, *Geophysics*, 2, 33.
 MANDELSTAM, L., 1938, *Izv. Akad. Nauk. S.S.S.R.*, 525.
 MEAD, L., R.R.D.E., Ministry of Supply, Research Report 268.
 PAPALEXI, N., 1938, *Izv. Akad. Nauk. S.S.S.R.*, 539.
 VARLEY, G. C., A.O.R.G. Memorandum No. 467.

The Application of a Variational Method to the Calculation of Radio Wave Propagation Curves for an Arbitrary Refractive Index Profile in the Atmosphere

By G. G. MACFARLANE

Telecommunications Research Establishment, Ministry of Supply, Gt. Malvern

MS. received 23 October 1947

ABSTRACT. The problem of determining the field strength produced by a microwave transmitter beyond the horizon has been intensively studied in recent years. Charts and tables, which are based on approximate solutions of the wave equation obtained by the phase-integral method, have been produced (Booker and Walkinshaw 1946), from which the field strength can be calculated when the modified refractive index of the air (the M -curve) varies with height according to a power law. In many cases, however, the M -curve cannot be adequately represented by a power law, and serious analytical difficulties are encountered if the problem is tackled by the phase-integral method.

The basic problem is to find the eigenvalues D_m and eigenfunctions U_m of the wave equation $d^2 U_m/ds^2 + \{s + f(s) + D_m\} U_m = 0$, in which $f(s)$, the M -anomaly, is a given function, which tends to zero as s tends to infinity. In this paper a method, previously described by the author (Macfarlane 1947), is applied to determine the first eigenvalue D_1 for the class of M -anomalies that can be represented by a series of the form, $\sum_{n=1}^r A_n \exp(-\alpha_n s)$.

Two numerical examples are given to illustrate the method. They were obtained by fitting one exponential and three exponentials in turn to an M -anomaly calculated from meteorological observations made at Kaikoura, New Zealand. In the first case the M -anomaly is represented by a surface duct and in the second case by an elevated duct. The exponential attenuation rate, which is determined by the imaginary part of D_1 , is calculated in both cases as a function of the wavelength.

Finally, the theoretical height-gain function for a wavelength of 3 metres is calculated from the M -curve observed 47 miles off shore from Ashburton, South Island, New Zealand, in the late afternoon of 4th November 1946. It is found that there is close agreement between theoretical and observed height-gain curves.

§ 1. INTRODUCTION

THE theory of propagation in current use for describing the field beyond the horizon of a radio transmitter expresses the field as a sum of normal modes of the wave equation. The independent variables of the wave equation, which are height and range, are separable, since the atmosphere is assumed to be horizontally stratified. This simplification reduces the problem of solving the wave equation with two independent variables to that of finding the eigenvalues and normal modes of the one-dimensional wave equation *

$$d^2 U_m/dh^2 + k^2 \{y(h) + \Lambda_m\} U_m = 0. \quad \dots\dots(1)$$

Each mode must satisfy the boundary conditions, that $U_m(0) = 0$, and that as $h \rightarrow \infty$ U_m represents an upgoing wave. In this equation

$$y(h) = N^2(h) - 1 = 2 \times 10^{-6} M(h), \quad \dots\dots(2)$$

where $N(h)$ is the modified refractive index of the air and Λ_m is the m th eigenvalue.

* Since the subject matter of this paper is closely related to that of Pekeris (1946) we shall adopt a similar notation.

A number of different methods have been used to evaluate U_m and Λ_m . Booker and Walkinshaw (1946) applied the phase-integral method, originally discussed in this connection by Eckersley and Millington (1938), to find U_m and Λ_m for the M -curve,

$$M(h) = K_\infty(h - d^{1-m}h^m/m) \times 10^{-6}, \quad \dots\dots(3)$$

where $m=0.5$. Here d is the duct width and K_∞ is the gradient of modified refractive index at infinite height.

The phase-integral method leads to solutions which are only approximately correct but which, nevertheless, may be sufficiently accurate for most practical purposes; the approximation that the atmosphere is a horizontally stratified medium is indeed less satisfactory than the analytical approximation made in the phase-integral theory. However, the phase-integral method leads to troublesome mathematical problems when the M -curve has a form more complicated than that of equation (3). Even when $m=0.2$ it was too complicated to evaluate analytically and it was found necessary to compute solutions by means of a differential analyser. Techniques for evaluating the required solutions of the wave equation on the differential analyser were devised by Hartree, Michel and Nicolson (1946).

Another method of evaluating U_m and Λ_m has been described in a recent paper by Pekeris (1946). The M -curve is expressed in the form

$$M(h) = K_\infty h + g\epsilon^{-ch}$$

and U_m and Λ_m are found by a perturbation method. The procedure is to express U_m as a linear combination of the eigenfunctions U_n^0 of all the modes in the standard case, in which $M(h) = K_\infty h$. The coefficients A_{mn} of U_n^0 and the eigenvalues Λ_m are found as solutions of an infinite system of equations. They are evaluated by an iterative procedure, which has been found to be rapidly convergent. Pekeris has also shown that a more general class of M -curves can be handled by a slight extension of the perturbation method. It includes M -curves of the forms

$$M(h) = K_\infty h + \sum_n g_n \exp(-c_n h) \quad \dots\dots(4.1)$$

$$\text{and} \quad M(h) = K_\infty h + \sum_n g_n h^n \exp(-c_n h). \quad \dots\dots(4.2)$$

Since an arbitrary M -curve can be represented by either of those series to any desired degree of accuracy, it might be said that the general problem has been solved. On the other hand the convergence of the procedure is least rapid when the mode is slightly leaky and this is a case of great interest.

The procedure we propose is based on a variational method in which the eigenvalue is found from the solution of a transcendental equation. It is not easy to estimate the accuracy of the eigenvalues obtained, but the eigenvalues found by it for the power-law M -curve of equation (3) agreed within 0.4% with those obtained by Hartree, Michel and Nicolson on the differential analyser (Macfarlane 1947).

We have developed the method for an M -curve of the form

$$M(h) = K_\infty h + g_0 + \sum_{n=1}^r g_n \exp(-c_n h), \quad \dots\dots(4.3)$$

so that it can be applied to any M -curve. If it is desired, an M -curve expressed in the form of equation (4.2) can also be handled with the information given in this paper.

Following Pekeris (1946) we introduce natural units of height

$$\left. \begin{aligned} s &= h/H & H &= (k^2 q)^{-\frac{1}{2}} \\ q &= dN^2/dh = 2K_{\infty} \times 10^{-6} = 2.5 \times 10^{-9} \text{ cm}^{-1} * \\ D_m &= \Lambda_m (k/q)^{\frac{1}{2}}, \end{aligned} \right\} \dots\dots(5)$$

so that equation (1) assumes the form

$$d^2 U_m / ds^2 + \{s + f(s) + D_m\} U_m = 0. \dots\dots(6)$$

In terms of s the M -curve, equation (4), gives

$$f(s) = A_0 + \sum_{n=1}^r A_n \exp(-\alpha_n s), \dots\dots(7)$$

where

$$A_n = w g_n, \dots\dots(7.1)$$

$$\left. \begin{aligned} \alpha_n &= 2 \times 10^{-6} c_n / w q, \\ &= 26.25 c_n / w, \quad c_n \text{ in ft}^{-1}, \end{aligned} \right\} \dots\dots(7.2)$$

and

$$\left. \begin{aligned} w &= 2 \times 10^{-6} (k/q)^{\frac{1}{2}} \\ &= 3.696 \lambda^{-\frac{1}{2}}, \quad \lambda \text{ in cm.} \end{aligned} \right\} \dots\dots(7.3)$$

In the work to be reported here we have confined our attention to the determination of the first eigenvalue D_1 and its eigenfunction $U_1(s)$. This has been done because, in many cases of importance, the contributions of higher modes to the field strength beyond the horizon is negligible. In terms of natural units of horizontal distance d ,

$$x = d/L, \quad L = 2(kq^2)^{-\frac{1}{2}}, \dots\dots(8)$$

and the field strength at range x and height s_2 due to a transmitter of power gain G and power P at height s_1 is

$$E = \frac{(120\pi PG)^{\frac{1}{2}}}{L} x^{-\frac{1}{2}} \sum_m \exp(iD_m x) U_m(s_1) U_m(s_2) \dots\dots(9.1)$$

$$\simeq \frac{(120\pi PG)^{\frac{1}{2}}}{L} x^{-\frac{1}{2}} \exp(iD_1 x) U_1(s_1) U_1(s_2) \dots\dots(9.2)$$

at sufficiently great range beyond the horizon. In these formulae the eigenfunction U_m is normalized so that

$$\int_0^{\infty} U_m^2(s) ds = 1. \dots\dots(10)$$

From equations (7.3) and (8) the numerical distance is expressible in terms of the wavelength parameter w as

$$x = 0.1423 w^{\frac{1}{2}} d, \quad d \text{ in miles.} \dots\dots(11)$$

The exponential attenuation rate of the first mode is, therefore,

$$8.686 x d^{-1} \text{Im}(D_1) = 1.236 w^{\frac{1}{2}} \text{Im}(D_1) \text{ db. per mile.} \dots\dots(12)$$

§ 2. THE VARIATIONAL METHOD

The variational equations of the wave equation, equation (6), in which the function $f(s)$ is given by equation (7), are (Macfarlane 1947)

$$D(\phi) = \int_0^{\infty} \left(\frac{d\phi}{dy} \right)^2 dy - \int_0^{\infty} y \phi^2 dy - \sum_{n=1}^r A_n \int_0^{\infty} \exp(-\alpha_n y) \phi^2 dy \dots\dots(13.1)$$

and

$$H(\phi) = \int_0^{\infty} \phi^2 dy. \dots\dots(13.2)$$

* We take the "effective" radius of the earth to be 8×10^6 metres.

It should be noticed that the constant term A_0 is absorbed into D_1 . $D_1 + A_0$ is then the minimum stationary value of the ratio D/H for functions ϕ that satisfy the boundary conditions (i) $\phi(0)=0$; (ii) $\phi(y) \rightarrow 0$ as $y \rightarrow \infty \epsilon$, where $-\pi/3 < \arg \epsilon < 0$, and that have no zero values on the contour from $y=0$ to $y=\infty \epsilon$. As explained by Eckersley and Millington (1938) a suitable choice for ϕ is the Airy function

$$\phi = \text{Ai}(cy - j_1), \quad \dots\dots(14)$$

where $-j_1 = -2.3381$ is the first zero.

Parameter c is determined so that the ratio D/H has a minimum stationary value and $\epsilon = 1/c$.

Substitution of function ϕ , given by equation (14), into the variational equations (13) gives

$$\frac{D(\phi)}{H(\phi)} = \frac{1}{Q} \left[P \left(c^2 - \frac{2}{c} \right) - \sum_{n=1}^r A_n F(\alpha_n/c) \right], \quad \dots\dots(15)$$

where

$$P = \int_0^\infty \{\text{Ai}'(y - j_1)\}^2 dy = 0.38321; \quad Q = \int_0^\infty \{\text{Ai}(y - j_1)\}^2 dy = 0.49170 \quad \dots\dots(16)$$

$$F(z) = \int_0^\infty \exp(-zy) \{\text{Ai}(y - j_1)\}^2 dy.$$

The stationary values of the ratio D/H occur at the saddle points of the function D/H in the complex c -plane; that is, where

$$\frac{d}{dc} \left(\frac{D}{H} \right) = \frac{1}{Q} \left[2P \left(c + \frac{1}{c^2} \right) + \frac{1}{c^2} \sum_{n=1}^r A_n \alpha_n F'(\alpha_n/c) \right] = 0$$

or

$$0 = c^3 + 1 + 1.3048 \sum_{n=1}^r A_n \alpha_n F'(\alpha_n/c). \quad \dots\dots(17)$$

The root of equation (17) lying in the fourth quadrant which gives the smallest value for the modulus of D/H , is the required value of c . On substitution of this value of c into equation (15) the approximate value of $D_1 + A_0$ is obtained. It should be noted that the constant term A_0 affects only the phase of the first mode, since, by equation (12), the attenuation rate is dependent solely on the imaginary part of D_1 .

The application of the variational method to the calculation of D_1 for an arbitrary M -curve involves therefore four steps:

(i) representation of the given M -curve by an expression of the form

$$M(h) = K_\alpha h + g_0 + \sum_{n=1}^r g_n \exp(-c_n h) \quad \dots\dots(18)$$

with r , the number of exponential terms, as small as possible;

(ii) calculation of the Laplace transform of the square of the Airy function and of its derivative for complex values of the argument z ,

$$F(z) = \int_0^\infty \exp(-zy) \{\text{Ai}(y - j_1)\}^2 dy \quad \dots\dots(19)$$

and

$$F'(z) = - \int_0^\infty y \exp(-zy) \{\text{Ai}(y - j_1)\}^2 dy; \quad \dots\dots(20)$$

(iii) solution of the minimal variation equation (17) to find c ;

(iv) calculation of $A_0 + D_1 = D/H$ from equation (15) using the values of c found from (iii).

When D_1 has been found the corresponding eigenfunction $U_1(s)$ is obtained by numerical integration or by use of a differential analyser.

For (i) one method of fitting curves of the form of equation (18) to given M -curves depends on putting $c_n = nc$ and rewriting equation (18) in the form

$$M(h) - K_\infty h - g_0 = \sum_{n=1}^r g_n \exp(-nch) = \sum_{n=1}^r g_n z^n, \text{ where } z = \exp(-ch).$$

The range of the variable z extends from 0 to 1. Since K_∞ and g_0 can be obtained from the asymptotic behaviour of $M(h)$ for large values of h , the problem becomes one of fitting a polynomial to a given set of points, in the range $0 \leq z \leq 1$. Gram polynomials, which have been extensively tabulated by Davis (1935), are admirably suited to this purpose.

For (ii) the functions $F(z)$ and $F'(z)$ have been calculated to five figures at intervals of 0.2 in the real and imaginary parts of z in the range $x = 0.0(0.2)4.0$, $y = 0.0(0.2)3.2$, where $z = x + iy$. Both real and imaginary parts are given in a paper by Hay (1947) together with coupled differences and asymptotic formulae for extending the range of the variables.

§ 3. NUMERICAL EXAMPLES

In order to illustrate the technique we shall consider the case of an M -curve calculated from meteorological observations made at 1941 hours on 10 October

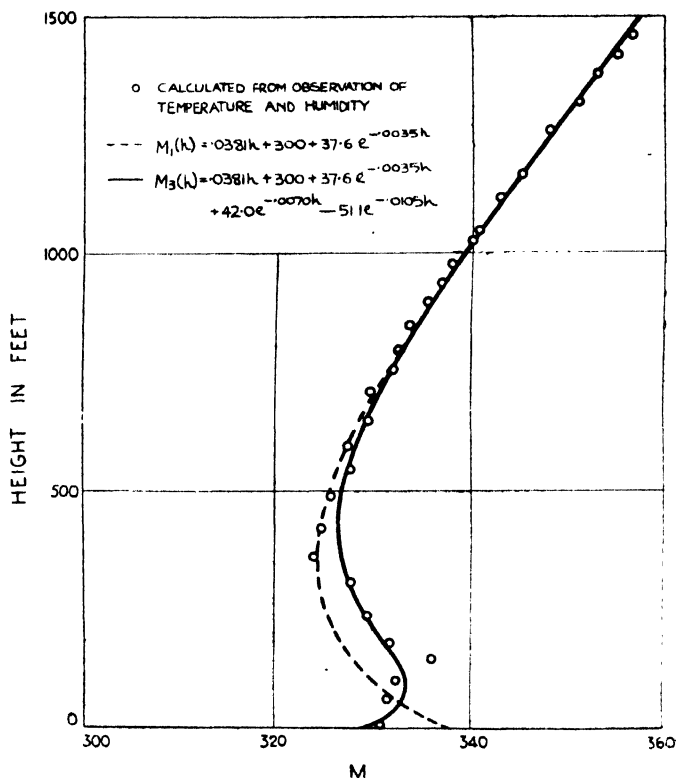


Figure 1. M -curve found at 1941 hrs. on 10 October 1944 at Kaikoura, New Zealand.

1944 at Kaikoura, New Zealand. This profile of modified refractive index can be represented by the following cubic function of exponentials *

$$M(h) = 0.0381h + 300 + 37.4e^{-0.0035h} + 42.0e^{-0.0070h} - 51.1e^{-0.0105h}. \quad \dots\dots(21)$$

* The author is indebted to Mr. J. W. Head for this formula.

The original M -curve and the approximation represented by equation (19) are shown in figure 1. In the same figure the approximation using only the first exponential term in equation (21) is also shown. In figure 2 the M -excess, denoted by μ and equal to $M(h) - 0.0381h - 300$, is plotted together with the approximations represented by one and three exponential terms. This shows more clearly than figure 1 the degree of fit achieved. An even closer fit using five exponential terms is also shown in figure 2, but we shall not make use of it.

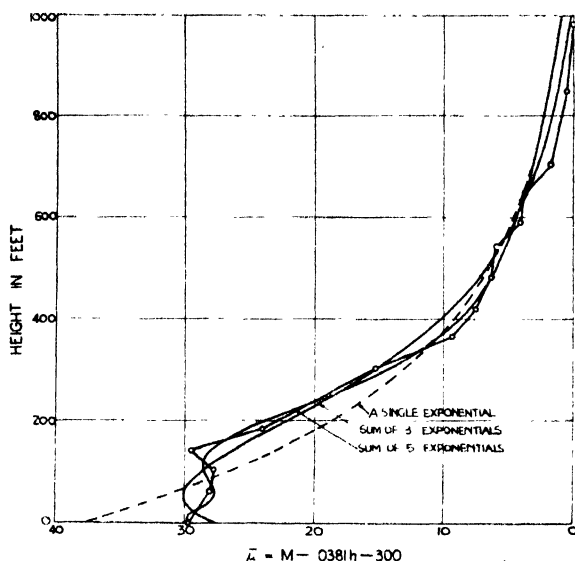


Figure 2. M -excess for M -curve of figure 1.

Figure 1 shows how one exponential term can represent a surface duct and how the addition of other two exponential terms can give rise to an elevated duct. As it is interesting to compare the eigenvalues for these two cases we shall illustrate our technique by working out both.

The first step is to find the constants A_n and α_n of the series for $f(s)$. These are given in terms of the wavelength parameter w by equations (7.1) and (7.2). Thus

$$\left. \begin{aligned} A_0 &= 300w \\ A_1 &= 37.4w & \alpha_1 &= 0.0918/w & A_1\alpha_1 &= 3.433 \\ A_2 &= 42.0w & \alpha_2 &= 0.1836/w & A_2\alpha_2 &= 7.711 \\ A_3 &= -51.1w & \alpha_3 &= 0.2754/w & A_3\alpha_3 &= -14.073 \end{aligned} \right\} \dots\dots (22)$$

We now consider separately the cases of the surface duct and the elevated duct.

(i) The surface duct

For the surface duct

$$M(h) = 0.0381h + 300 + 37.4e^{-0.0085h} \dots\dots (23)$$

and A_1 and α_1 have the values given in equation (22).

Equation (17), which determines c , is

$$c^3 + 1 + 4.479F'(0.0918/cw) = 0. \dots\dots (24)$$

To solve equation (24) put

$$z = 0.0918/cw \dots\dots (25)$$

and rewrite it as $(0.0918/w)^3 = -z^3\{1 + 4.479F'(z)\} \dots\dots (26)$

Since w is entirely real the problem is to find values of z which make the imaginary part of the right-hand side of equation (26) zero. Moreover, since $\arg z = -\arg c$ and since only values of c are allowed for which $0 < \arg c < \pi/3$ we need consider only values of z for which $-\pi/3 < \arg z < 0$. It is best to consider trapped modes, for which w , c and z are all real, separately from leaky modes, for which c and z are complex.

Next the exponential attenuation rate $1.236w^{\frac{1}{2}}\text{Im}(D_1)$ is calculated for each value of w . Finally the wavelength corresponding to each value of w is read off figure 4, which is a plot of the function $3.696\lambda^{-1}$.

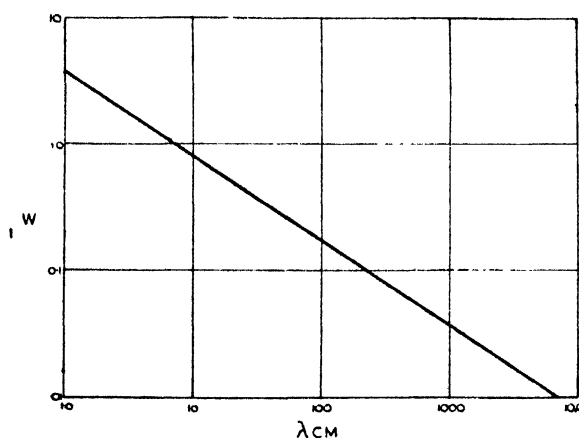


Figure 4. Parameter w as function of wavelength.

Table 2 gives the final results.

Table 2. The eigenvalues and the exponential rates as functions of the wavelength for the dotted M -curve of figure 1

z		w		c		$300w + D_1 = D/H$		γ	λ (cm.)
Re	Im		Re	Im	Re	Im			
0.0	0.0	∞	1.335	0.0	$-\infty$	0.0	0.0	0.0	0.0
0.2	0.0	0.4117	1.115	0.0	-11.87	0.0	0.0	0.0	26.5
0.3	0.0	0.3057	1.001	0.0	-8.13	0.0	0.0	0.0	42
0.4	0.0	0.2597	0.884	0.0	-6.59	0.0	0.0	0.0	54
0.518	0.0	0.2422	0.732	0.0	-6.04	0.0	0.0	0.0	59.6
0.5512	-0.2	0.2008	0.7329	0.2659	-4.795	0.100	0.055	79	
0.6280	-0.4	0.1412	0.7365	0.4691	-3.094	0.456	0.212	134	
0.7180	-0.6	0.1034	0.7284	0.6087	-2.111	0.835	0.332	213	
0.8067	-0.8	0.0803	0.7144	0.7085	-1.594	1.148	0.402	311	
0.9680	-1.2	0.0561	0.6664	0.8262	-1.198	1.556	0.456	533	
1.1139	-1.6	0.0438	0.6177	0.8758	-1.178	1.775	0.459	775	
1.262	-2.0	0.0364	0.5694	0.9026	-1.230	1.896	0.447	1022	
∞	$-\infty$	0.0	0.5000	0.8660	-1.169	2.025	0.0	∞	

γ = exponential attenuation rate $1.236w^{\frac{1}{2}}\text{Im}(D_1)$

We shall now consider the better approximation to the experimental values of M that is given by three exponential terms.

(ii) The elevated duct

For the elevated duct

$$M(h) = 0.0381h + 300 + 37.4e^{-0.0035h} + 42.0e^{-0.0070h} - 51.1e^{-0.0105h}. \quad \dots\dots(28)$$

In place of equation (26) we now get

$$(0.0918/w)^3 = -z^3\{1 + 4.479F'(z) + 10.061F'(2z) - 18.362F'(3z)\}. \quad \dots\dots(29)$$

As before, we have to find values of z that make the imaginary part of the right-hand side of equation (29) zero. The procedure is the same as in (i) above and the final result is given in table 3.

Table 3. The eigenvalues and the exponential attenuation rates as functions of the wavelength for the continuous M -curve of figure 1

z		w	c		$300w + D_1 = D/H$		γ	λ (cm.)
Re	Im		Re	Im	Re	Im		
0.177	0.0	∞	0.0	0.0	$-\infty$	0.0	0.0	0.0
0.2	0.0	0.8085	0.5677	0.0	-26.28	0.0	0.0	9.8
0.3	0.0	0.3663	0.8354	0.0	-11.20	0.0	0.0	32.0
0.4	0.0	0.2669	0.8600	0.0	-7.692	0.0	0.0	51.5
0.5	0.0	0.2254	0.8146	0.0	-6.490	0.0	0.0	66.4
0.6	0.0	0.2089	0.7324	0.0	-5.672	0.0	0.0	74.4
0.625	0.0	0.2084	0.705	0.0	-5.648	0.0	0.0	74.7
0.6586	-0.2	0.1748	0.7300	0.2217	-4.558	0.075	0.039	97.2
0.723	-0.4	0.1275	0.762	0.422	-3.036	0.369	0.163	156
0.7956	-0.6	0.0945	0.779	0.587	-2.040	0.751	0.285	244
0.867	-0.8	0.0749	0.764	0.705	-1.479	1.081	0.366	347
0.930	-1.0	0.0626	0.732	0.786	-1.280	1.332	0.412	440
1.107	-1.6	0.0430	0.624	0.902	-1.178	1.742	0.447	790
∞	$-\infty$	0.0	0.500	0.866	-1.169	2.025	0.0	∞

$\gamma = \text{exponential attenuation rate } 1.236w \frac{1}{\lambda} \text{Im}(D_1).$

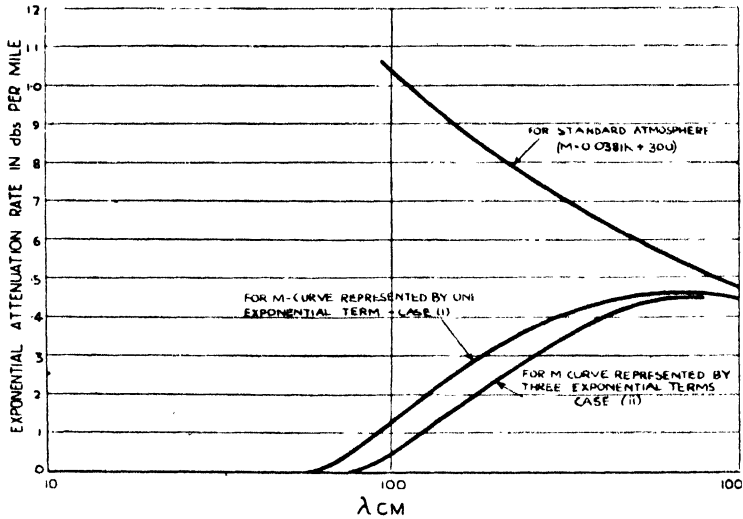


Figure 5. Dependence of exponential attenuation rate on wavelength.

Finally, in figure 5 we have shown the dependence of exponential attenuation rate on wavelength for the approximations to the experimental M -curve of figure 1 given by the surface and the elevated ducts.

The curve representing a standard atmosphere is also shown. It is seen that the attenuation rate is greater in case (i) than in case (ii). For $\lambda > 10$ metres

there is very little difference between all three curves but for $\lambda < 6$ metres the attenuation rate for the standard atmosphere is much higher than for the atmosphere represented by the curves of figure 1. In considering the curves of figure 5 it should not be forgotten that the total attenuation includes that due to the factor $x^{-\frac{1}{2}}$ of equation (9.2) as well as that due to the exponential attenuation rate, which is shown.

It must be emphasized that cases (i) and (ii) have been considered in detail primarily to illustrate the power and simplicity of this method for determining eigenvalues. In some cases as many as seven exponential terms might be required to represent adequately the experimental M -curve, but even so the extra labour involved is small.

§ 4. COMPARISON OF THEORETICAL AND EXPERIMENTAL HEIGHT-GAIN CURVES

The observed M -curve, shown by dots in figure 6, and the height-gain curve, shown in figure 7, were observed 47 miles off shore from Ashburton, South Island, New Zealand, in the late afternoon of 4 November 1946. The radio field was produced by a set operating on a wavelength of 3 metres at a height of 35 feet at Ashburton. Meteorological ascents made at different points in the path of transmission showed that the atmosphere was very nearly horizontally stratified, so that this assumption in the theory was justified.

The M -curve represents an interesting situation in which a

duct just fails to occur; M is constant up to about 400 feet and then increases linearly. For this M -curve only the first mode should be important at a range of 47 miles below the horizon. The situation was therefore simple enough for single-mode wave theory to apply. The method of analysis follows the lines described above: the M -curve is represented by a linear term modified by two exponential terms: $M(h) = 310 + 0.0365h + 46.4e^{-0.0046h} - 27.8e^{-0.0089h}$. This function is plotted in figure 6 together with the experimental points, and it will be seen that it represents a satisfactory approximation. The eigenvalue D_1 is then calculated for a range of wavelengths (table 4).

Lastly, the height-gain function is calculated by numerical integration of the wave equation using the eigenvalue appropriate to $\lambda = 3$ metres (table 5).

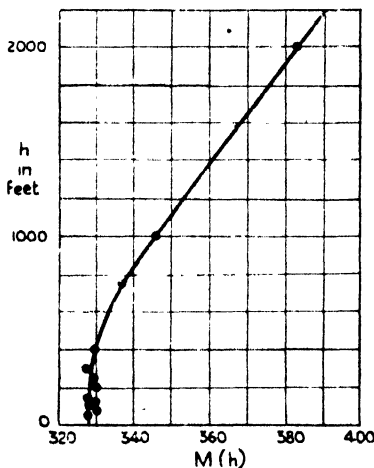


Figure 6. M -curve found 47 miles offshore from Ashburton, South Island, New Zealand, in late afternoon 4 November 1946.

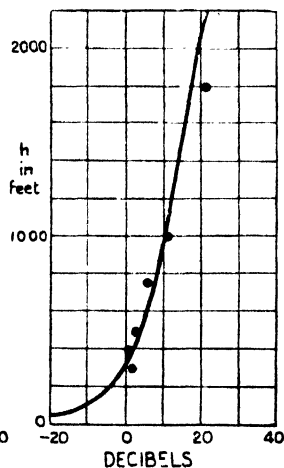


Figure 7. Height gain on $\lambda = 3$ m. The dots represent measured values; the full line is obtained from single-mode wave theory.

Table 4. The eigenvalues and the exponential attenuation rates as functions of the wavelength for the *M*-curve of figure 6

α		w		c		$310w + D_1 = D/H$		γ	λ (cm.)
Re	Im			Re	Im	Re	Im		
0.4047	-0.2	0.5796	0.4134	0.2040	-10.933	0.152	0.130	16.1	
0.5685	-0.4	0.2488	0.5710	0.4018	-4.505	0.489	0.302	57.3	
0.7106	-0.6	0.1551	0.6396	0.5401	-2.666	0.817	0.398	116.3	
0.8319	-0.8	0.1130	0.6674	0.6418	-1.893	1.087	0.451	187.1	
0.9360	-1.0	0.0895	0.6731	0.7191	-1.507	1.309	0.484	265.4	
1.0268	-1.2	0.0747	0.6654	0.7776	-1.301	1.481	0.500	348.1	
1.1077	-1.4	0.0645	0.6507	0.8226	-1.190	1.615	0.507	433.8	
					-1.402	1.387	0.492	300	

γ =exponential attenuation rate in db/mile.

Table 5. Theoretical height-gain function

h (feet)	159	318	445	636	764	954	1080	1272	1400	1590
$20 \log U_1 $	-6.07	-0.17	2.72	5.87	7.57	9.77	11.09	12.92	14.03	15.62

Figure 7 shows a plot of the theoretical and experimental height-gain functions. It will be seen that the agreement is good.

It is also interesting to consider how the exponential attenuation rate varies with wavelength. The results already given in table 4 are plotted in figure 8 with those for a standard atmosphere and for the 350-foot surface duct of § 3(i). It will be seen that the attenuation rate is small, but not zero, even at very short wavelengths, which is accounted for by the fact that the *M*-curve fails to show any duct. This type of *M*-curve, with a vertical tangent at ground level and no duct, represents a limiting case which divides *M*-curves into two categories: those with ducts and those without ducts.

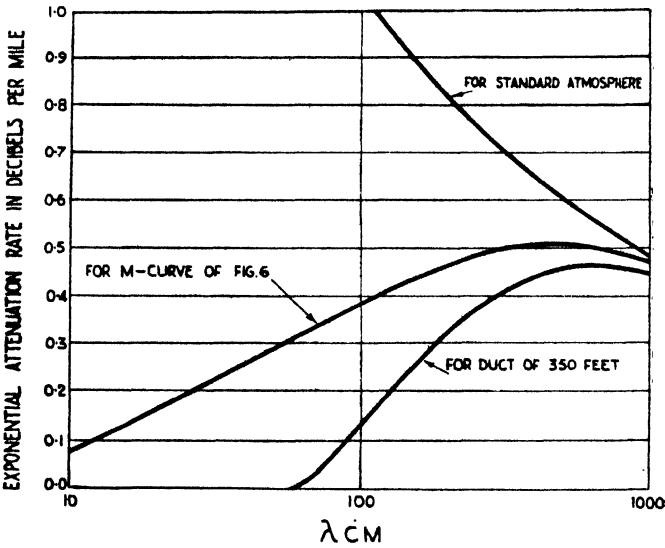


Figure 8. Dependence of exponential attenuation rate on wavelength for the *M*-curves of figures 1 and 2.

ACKNOWLEDGMENT

This work has been done at T.R.E., a Ministry of Supply Establishment, under contract to D.S.I.R. and as part of the Radio Research Board (D.S.I.R.) programme. It has been carried out in conjunction with D.S.I.R. New Zealand,

to whom thanks are due for permission to use the radio and meteorological data recorded at Ashburton, New Zealand, on 4 November 1946.

Crown copyright is reserved. The paper is reproduced with the permission of the Controller of H.M. Stationery Office.

REFERENCES

- BOOKER, H. G., and WALKINSHAW, W., 1946, *Meteorological Factors in Radio-Wave Propagation* (London: Physical Society), p. 80.
DAVIS, H. T., 1935, *Tables of the Higher Mathematical Functions*, Vol. II (Indiana: Principia Press Inc.).
ECKERSLEY, T. L., and MILLINGTON, G., 1938, *Phil. Trans. Roy. Soc. A*, **237**, 273.
HARTREE, D. R., MICHEL, J. G. L., and NICOLSON, P., 1946, *Meteorological Factors in Radio-Wave Propagation* (London: Physical Society), p. 127.
HAY, H. G., 1948, *Phil. Mag.*, in press.
MACFARLANE, G. G., 1947, *Proc. Camb. Phil. Soc.*, **43**, 213.
PEKERIS, C. L., 1946, *J. Appl. Phys.*, **17**, 678.

Radio Shadow Effects Produced in the Atmosphere by Inversions

By W. L. PRICE

Division of Radiophysics, Council for Scientific and Industrial Research,
Commonwealth of Australia

MS. communicated by L. G. Dobbie; received 21 July 1947

ABSTRACT. Reports from radar operators in Australia in 1940 revealed marked increase in range of surface objects on certain occasions apparently associated with distinctive meteorological conditions. It appeared likely that this phenomenon should be accompanied by the existence of "shadow" zones in the atmosphere, in which aircraft would not be detected. A theoretical investigation by ray-tracing methods shows that such shadow zones should exist above certain inversions. Controlled experiments, in which measurements of all the relevant physical conditions were made, show that these shadow zones do exist. The position and extent of the zones agrees well with the theory.

§ 1. INTRODUCTION

SHORTLY after the installation of coastal radar equipment on the east coast of New South Wales, in May 1940, operators began to report the occasional detection of surface objects at abnormally great ranges. The Commander Fixed Defences, Major-General J. S. Whitelaw, pointed out to Dr. D. F. Martyn that this effect coincided usually with the existence of an extensive anticyclone system over the eastern part of the Australian continent. Theoretical and experimental investigation was initiated, at first by the Radiophysics Laboratory of the Council for Scientific and Industrial Research, and later by the Australian Army Operational Research Group (O.R.G.).

It was early realized that the phenomenon was mainly due to the sharp temperature and humidity gradients associated with certain meteorological conditions, notably temperature inversions (Englund, Crawford and Mumford 1938), and attention was concentrated on this aspect. Information soon reached

Australia that a similar investigation in England, initiated principally by Booker and Walkinshaw (1946), was directed mainly towards the determination of the increased fields close to the surface of the earth, in the "radio duct" formed within and below such gradients by the "trapping" of radiation. Now it seemed probable that under abnormal conditions of high surface fields there should occur a weakening of the field above the abnormal gradients. If sufficiently severe this might explain the occasional failure of aircraft warning radar sets to detect aircraft flying at certain heights when within the normal range of the set (Martyn 1943). Accordingly attention was devoted in Australia to this latter aspect, which was of considerable operational importance.

In the early stages of the investigation an attempt was made to derive information from the log books of radar operators and aircraft pilots. However, this method proved unreliable and a series of carefully controlled flights off the Sydney coast was made during the summer of 1942-3. Steep climbs were made by an aircraft at a range of about 24 miles from a 200 Mc/s. transmitter, 250 feet above mean sea level. Field strength and temperature were measured in the plane. Humidity was not measured, but use was made of hygrometer readings taken by the R.A.A.F. during their regular meteorological flights at Richmond, about 25 miles inland. At the same time propagation through a horizontally stratified medium was investigated theoretically by a geometrical optics treatment, this being considered best for the region near the level of inversion (Price 1943). The experimental results were published in a service report by D. R. Brown (1943). Examination of the results showed that on days of temperature inversion the three lowest lobes of radiation were depressed from their normal position, while higher lobes were not measurably affected. Analysis by ray methods showed that the extent of the lowering was of the right order (Price 1943). The results indicated that further experimental work was desirable, especially along the lines of combined air sounding and field strength measurements. Work was commenced on the development of accurate instruments for this purpose, and taken over by the Radiophysics Laboratory after the Operational Research Group was disbanded in 1944.

Upon the completion of sounding equipment designed for use in aircraft, a series of flights was carried out off the N.S.W. coast near Sydney in January 1946 by the Radiophysics Laboratory. The results proved of interest in connection with (a) the development of inclined stratified air masses produced by off-shore streaming of hot dry air, modified to some extent by a sea-breeze type of circulation, and (b) the sharp decrease of field strength with height experienced on passing through a layer with negative refractive index gradient.

This paper discusses the ray theory of propagation of ultra-high-frequency radio waves through such layers. The general theory for inclined layers developed in § 2 is applied in §§ 3 and 4 to an analysis of the field above the region of surface trapping caused by a negative refractive index gradient. Rays have been traced for one of the January flights and the results are compared in § 5 with radar observations of echo strength from the aircraft.

§ 2. RAY THEORY FOR A STRATIFIED ATMOSPHERE

2.1. *Scope of ray theory*

An essential feature of ray theory is that the radiation field can be described in terms of the propagation of energy along rays whose direction is given by

the laws of refraction, a given parcel of electromagnetic flux travelling within a tubular space bounded by the rays. Surfaces of equal phase, orthogonal to the rays, constitute the wave front. This description will be valid, provided that the laws of the electromagnetic field (i.e. Maxwell's equations) are satisfied. It is known that there are limitations to the validity of ray theory, the main ones being:—

- (i) $|\text{grad } \mu| \ll 1/\lambda$, where λ is the wavelength. This means that the radius of curvature of the ray must be sufficiently greater than λ .
- (ii) The rays must not come to a focus.
- (iii) The region considered is not near the boundary rays produced by obstacles, nor within confined spaces whose dimensions are comparable with the wavelength.

To some extent ray theory can be extended by the use of Huyghens' Principle and the idea of interference. When this method becomes too complex other solutions of the wave equation are sometimes available which then enable the field to be described in simpler terms, such as, for example, by means of transverse electric (or magnetic) modes. Which is used depends on the nature of the problem and the completeness required in the solution. Treatment by ray methods, keeping their limitations in mind, sometimes provides a useful if approximate solution when a more complete quantitative answer is arithmetically, if not algebraically, impracticable. This appears to be the case when dealing with the propagation of U.H.F. radio waves over the earth, just above a refractive index inversion, in the region where the lobe pattern begins to be sensibly affected by super-refraction.

2.2. Equation of the ray

Figure 1 shows a ray TR from a transmitter T, at height h_T above the surface of the earth, which is assumed spherical with radius a , and centre O. The position of any point R is given by polar coordinates $(a+h, \theta)$ or by height h and reduced range d measured along the surface of the earth. The contours of equal refractive index, $\mu = \text{constant}$, may be curved with respect to the earth and inclined at small angle to the horizontal. For simplicity, cylindrical symmetry about the vertical through the transmitter is assumed, so that the path of a ray lies in a vertical plane. In general, this would not be strictly accurate and lateral deviations (which could be important in bearing measurements) are possible, but since all deviations are small, it is considered that vertical and horizontal refraction can be treated independently.

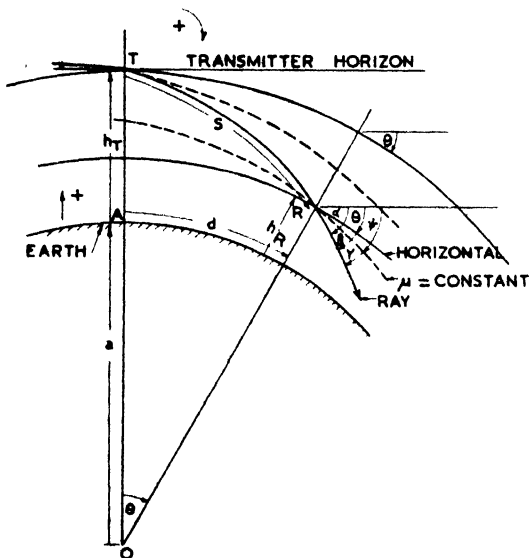


Figure 1. Coordinate system used in text for a ray passing through inclined layers.

from P to R; $\delta_2\gamma$, equal to the angular displacement at R of the first surface relative to the second. $\delta\psi$ may be similarly subdivided. Then

$$\delta\gamma = \delta_1\gamma + \delta_2\gamma, \quad \delta\psi = \delta_1\psi + \delta_2\psi, \quad \delta_2\gamma = -\delta_2\psi, \quad \delta\alpha = \delta\gamma + \delta\psi = \delta_1\gamma + \delta_1\psi. \quad \dots\dots(2)$$

$\delta_2\psi$ is due to the displacement of the centre of curvature.

Equation (1) may be written

$$\cos\gamma \frac{\delta\mu}{\mu\delta_1n} = \sin(\gamma + \delta_1\gamma/2) \frac{\delta\alpha}{\delta_1n} + \{\sin(\gamma + \delta\gamma/2) - \sin(\gamma + \delta_1\gamma/2)\} \frac{\delta\alpha}{\delta_1n},$$

$$\text{i.e.} \quad \frac{\cos\gamma(\delta\mu/\mu + \delta_2\psi \cdot \delta\alpha/2)}{\delta_1n} \simeq \sin(\gamma + \delta_1\gamma/2) \frac{\delta\alpha}{\delta_1n} = -1/\rho. \quad \dots\dots(3)$$

since

$$\cos(\gamma + \delta\gamma/4 + \delta_1\gamma/4) \simeq \cos\gamma,$$

$1/\rho$ is here the average curvature of the ray PR.

Equation (3) may be rearranged in the following form, suitable for computation:

$$\cos\gamma \frac{\delta\mu}{\mu\delta_1n} \left(\frac{1}{1 - \cos\gamma(\delta_2\psi/2)/\sin(\gamma + \delta_1\gamma/2)} \right) = \sin(\gamma + \delta_1\gamma/2) \frac{\delta\alpha}{\delta_1n}. \quad \dots\dots(4)$$

From the geometry of PQ'R, to the first order of small quantities,

$$\cos\gamma/r_1 = -\sin(\gamma + \delta_1\gamma/2)\delta_1\psi/\delta_1n, \quad \dots\dots(5)$$

remembering that δn is positive when directed upwards and r_1 is positive when the surface is concave downward. From equations (4) and (5), noting that γ_1 can be put equal to γ , we get the following relations:

$$\cos\gamma_1 \left\{ \frac{\delta\mu}{\mu\delta_1n} \left(\frac{1}{1 - \cos\gamma_1(\delta_2\psi/2)/\sin(\gamma_1 + \delta_1\gamma/2)} \right) + \frac{1}{r_1} \right\} = \sin(\gamma_1 + \delta_1\gamma/2) \frac{\delta_1\gamma}{\delta_1n} \quad \dots\dots(6)$$

$$= \cos\gamma_1(1/r_1 - 1/\cos\gamma_1\rho) = \cos\gamma_1/k_1, \quad \dots\dots(7)$$

which is the average curvature of stratification at P relative to a ray inclined to it at γ . $1/k_1^*$ can be called the average "effective curvature" (Schelleng, Burrows and Ferrell 1933) of the stratification at P.

When $\gamma \ll 1$ and $\mu \simeq 1$ we get

$$\frac{\delta_1n}{k_1} = \delta\mu \left(\frac{1}{1 - (\delta_2\psi/2)/(\gamma_1 + \delta_1\gamma/2)} \right) + \frac{\delta_1n}{n_1} = (\gamma_1 + \delta_1\gamma/2)\delta_1\gamma = \delta_1(\gamma^2/2), \quad \dots\dots(8)$$

$(\gamma_1 + \delta_2\gamma/2)$ is the average slope of the ray relative to the first surface and is approximately equal to $-\delta_1n/\delta_1d$. $(\delta_2\psi/2)(\gamma_1 + \delta_1\gamma/2)$ can always be made small compared with unity by decreasing $\delta\mu$ sufficiently.

Since it can be shown that

$$k_1/r_1 = -\delta_1\psi/\delta_1\gamma \quad \dots\dots(9)$$

from (5), (6) and (7), it follows that the range measured along the stratification at P will be given by

$$\delta_1d = r_1\delta_1\psi = -k_1\delta_1\gamma. \quad \dots\dots(10)$$

* Sometimes the symbol k is used for the ratio ("effective radius"/actual earth's radius), instead of for effective radius, as here.

The optical distance is

$$(\mu + \delta\mu/2) \cos(\gamma_1 + \delta_1\gamma/2) \delta_1 d - (\mu + \delta\mu/2) \sin(\gamma_1 + \delta_1\gamma/2) \delta_1 n \quad \dots\dots (11)$$

and so can be separated into two components, one depending upon the distance along the stratum and the other along the normal. $-\delta_2\psi/\delta_1 n$, which is a measure of the curvature $1/q$ of RQ, the line orthogonal to the strata, and the curvatures $1/r_1$, $1/\rho$, must be finite or zero for all values of γ as $\delta\mu$, $\delta\alpha$ etc. approach zero, unless there is a discontinuity in the slope of the contours. Thus, in the limit, equations (6) and (7) reduce to

$$\frac{\cos \gamma}{k} = \cos \gamma \left(\frac{d\mu}{\mu dn} + \frac{1}{r} \right) = \sin \gamma \frac{d_1\gamma}{dn} = \sin \gamma \left(\frac{d\gamma}{dn} - \frac{1}{q} \right) \quad \dots\dots (12)$$

when $\delta\mu \rightarrow 0$. Also

$$1/k = (1/r + d\mu/\mu dn) = (1/r - 1/\rho_T), \quad \dots\dots (13)$$

$$d_1 d = r d_1 \psi = -k d_1 \gamma. \quad \dots\dots (14)$$

In the limit the effective curvature of the stratification, $1/k$, becomes equal to the curvature of this surface, relative to the tangent ray at P, whose curvature is

$$1/\rho_T = -(1/\mu)(d\mu/dn).$$

Numerical computation of the ray path. When γ is small equations (8) and (10), giving finite differences, can be used for tracing the path of the ray. If a numerical procedure is adopted the path of the ray is determined in steps between consecutive surfaces $\mu = \text{constant}$, spaced so that the curvatures of each surface and of the ray are approximately constant, and the changes in their slopes ($\delta\psi$, $\delta\alpha$) are small. $\delta_2 n$, $1/r_1$, $1/r_2$, γ_1 and $\{\delta_2\psi - (\delta d/r_2 - \delta d/r_1)\}$, quantities which occur at the beginning of a step, can be read at each stage from a large scale plot of the contours, similar to figure 8 (a). The last quantity is equal to the initial difference in ψ between the two surfaces. An approximate value of δd can be obtained by inspection of the previous path of the ray.

Values of $\delta_1 n$, $\delta_2\psi$ and $(\gamma + \delta_1\gamma/2)$ are then calculated from the relation

$$(\gamma + \delta_1\gamma/2) = -\delta_2 n / \delta d + \{\delta_2\psi - (\delta d/r_2 - \delta d/r_1) + \frac{1}{2}(\delta d/r_2 - \delta d/r_1)\} = -\delta_1 n / \delta d. \quad \dots\dots (15)$$

Using these values, (8) is solved for $\delta_1\gamma$ by successive approximation. With the new value of $(\gamma + \delta_1\gamma/2)$, (15) is solved for δd leading to new values of $\delta_2\psi$ and $\delta_1 n$. The process is repeated in (8) and (15) until consistent values of $\delta_1\gamma$ and $\delta_1 d$ are obtained. Equation (10) should then check automatically.

It is convenient to use the modified refractive index described in § 2.3.2.

2.3. Transformation of coordinates

The form of expressions (12), (13) and (14) shows that the relative curvature of the ray, at any point, viz. $-(\cos \gamma)/k$, is derivable from a "modified" refractive index N whose logarithmic gradient dN/Ndn is equal to $1/k$, the stratum at the point being regarded as flat. We may also regard the process as one of subtracting a curvature $(\cos \gamma)/r$ from any ray making an angle γ to the stratification. Inclination of the contours due to change in the position of the centre of curvature is not thereby eliminated.

2.3.1. *Concentric stratification.* When the strata are concentric (13) becomes

$$1/k = d\mu/\mu dr + 1/r = \tan \gamma d\gamma/dr, \quad \dots\dots (16)$$

since now $d\gamma = d_1\gamma$; $d_2\gamma = 0$; $d_1n = dr$. Integrating, we get the well known equation of the ray

$$\mu r \cos \gamma = \text{constant.} \quad \dots\dots(17)$$

2.3.2. *Horizontal stratification.* Further, if the stratification is horizontal

$$1/k = d\mu/(\mu dh) + 1/(a+h) = dN/(Nd h) = \tan \gamma d\gamma/dh \quad \dots\dots(18)$$

$$\simeq d(\gamma^2/2)/dh, \quad \gamma \ll 1,$$

and

$$dd = ad\psi = -akd\gamma/(a+h),$$

where d is here the range, reduced to the earth's surface.

Integrating (18) with $N = \mu^*$ when $h = 0$ we get

$$N - 1 = \mu(a+h)/a - 1 = (\mu - 1) + h/a + (\mu - 1)h/a \quad \dots\dots(19)$$

and

$$N \cos \gamma = \text{constant.} \quad \dots\dots(20)$$

The following approximation to the "modified" index N is generally used (Freehafer 1943):

$$N - 1 = (\mu - 1) + h/a, \quad \mu \simeq 1, \quad h \ll a. \quad \dots\dots(21)$$

Also it is customary to define a quantity M , of convenient size, by

$$10^6 M = (\mu - 1) + h/a. \quad \dots\dots(22)$$

2.3.3. *Plot in rectangular coordinates* (Booker and Walkinshaw 1946).

It follows from the last section that, in a horizontally stratified medium situated over a flat earth and having a refractive index N given by (18) or (19), a ray will have the same relation between γ , h and reduced range d , as in a horizontal curved medium, refractive index μ , over a curved earth of radius a . Thus a plot in rectangular coordinates of reduced range against height will give the same ray paths as would be obtained by straightening the strata and making their length equal to their projection to ground level. These will not be quite the same curves as obtained using N over a plane earth, unless the ranges in the latter case are also reduced by the factor $a/(a+h)$.

It is convenient to eliminate the curvature of the earth and use a modified index of refraction even in the case of non-horizontal stratification. When h/a is now added to μ to obtain N , the lines of constant N will not remain parallel to those of constant μ , but become inclined at an angle ϵ , given by

$$(-d\mu/d_1n) \sin \epsilon = (1/a) \sin (\psi - \theta + \epsilon), \quad \dots\dots(23)$$

this being the condition for zero gradient of N in the direction inclined at ϵ to the lines $\mu = \text{constant}$ (figure 3).

It will be shown that the path of the ray will be approximately the same when N is taken as the refractive index and the earth's curvature disregarded as when the actual refractive index and curvature are used, provided the angles of inclination are small.

In figure 3 the angles and radii are defined as before, dashes distinguishing those measured to

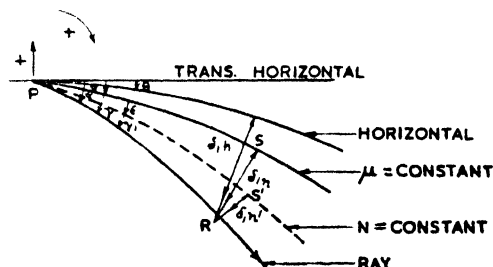


Figure 3. Transformation from contours of $\mu = \text{constant}$ to those of $N = \text{constant}$.

* The unnormalized value of N is here defined.

$N = \text{constant}$. Cosines of small angles can be put equal to 1. We have from equation (1)

$$\frac{d\mu}{d_1 n} = \tan \gamma \frac{d\alpha}{d_1 n} \quad (\mu \simeq 1) = \frac{\cos \gamma'}{\cos \gamma} \tan \gamma' \frac{d\alpha}{d_1 n'}$$

since angles RSP, RS'P each equals 90° .

$$\text{Therefore } \left(\frac{d\mu}{d_1 n} \cos \epsilon + \frac{\cos(\psi - \theta + \epsilon)}{a} \right) - \frac{\cos(\psi - \theta + \epsilon)}{a} = \cos \epsilon \frac{\cos \gamma'}{\cos \gamma} \tan \gamma' \frac{d\alpha}{d_1 n'}$$

$$\text{But } 1/r_N = -\tan \gamma' \frac{d_1(\psi + \epsilon)}{d_1 n'} = \text{curvature of PS'};$$

on putting $d\alpha = d(\psi + \epsilon + \gamma')$ and adding the last two equations we get

$$\left(\frac{d\mu}{d_1 n} \cos \epsilon + \frac{\cos(\psi - \theta + \epsilon)}{a} \right) + \left(\frac{1}{r_N} - \frac{\cos(\psi - \theta + \epsilon)}{a} \right) \simeq \tan \gamma' \frac{d_1 \gamma'}{d_1 n'}$$

$$\text{or } dN/d_1 n' + 1/r' = \tan \gamma' d_1 \gamma' / d_1 n' \quad \dots\dots (24)$$

since $1/r' = 1/r_N - \cos(\psi - \theta + \epsilon)/a = \text{curvature of the } N \text{ stratum relative to the earth}$ and $((d\mu/d_1 n) \cos \epsilon + \cos(\psi - \theta + \epsilon)/a)$ is the resultant gradient of N . Equation (24) gives the same form for the variation of γ with height as when N is the refractive index and the earth is flat.

§ 3. CHARACTERISTIC OF N AND M CURVES FOR HORIZONTAL STRATIFICATIONS

An advantage of the use of a modified index lies in showing quickly how the inclination of a ray to the horizontal changes with height. When the gradient of N (or M) is positive, γ continually decreases with height and the ray is convex upwards relative to the earth; conversely for a negative gradient. The usual condition of the atmosphere is for grad N to have a positive value lying between $30 \times 10^{-9}/\text{ft.}$ and $45 \times 10^{-9}/\text{ft.}$ and increasing with height. The standard value of $1/k$ generally used at low levels is $36 \times 10^{-9}/\text{ft.}$ Temperature inversions and, more especially, abnormal hydrolapses producing negative gradients of μ do occur quite frequently, but they can only persist over limited intervals of height. The resulting structure of the atmosphere shows a series of maxima and minima of N with height, the greatest and most important N deficits occurring below the first one or two thousand feet. Higher level N inversions which occur can reflect radiation to an extent which is sometimes significant, the amount depending on the shape of the (N, h) curve at the inversion, and on the propagation frequency. If only the lower inversions are considered, it is found that the (N, h) curves fall into several definite classes as follows:—

(a) *Figure 4.* The gradient of N is always positive. All rays are concave up, relative to the earth. Sub-standard propagation occurs when the value of the gradient of N is greater than normal ($36 \times 10^{-9}/\text{ft.}$).

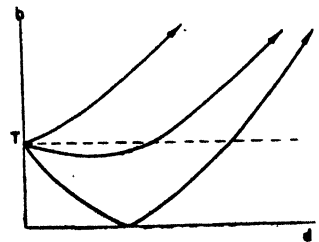


Figure 4. Paths of a ray when gradient M is positive.

(b) *Figure 5.* The initial gradient is negative and a minimum occurs at B. Rays are concave down or up, according as they are below or above B.

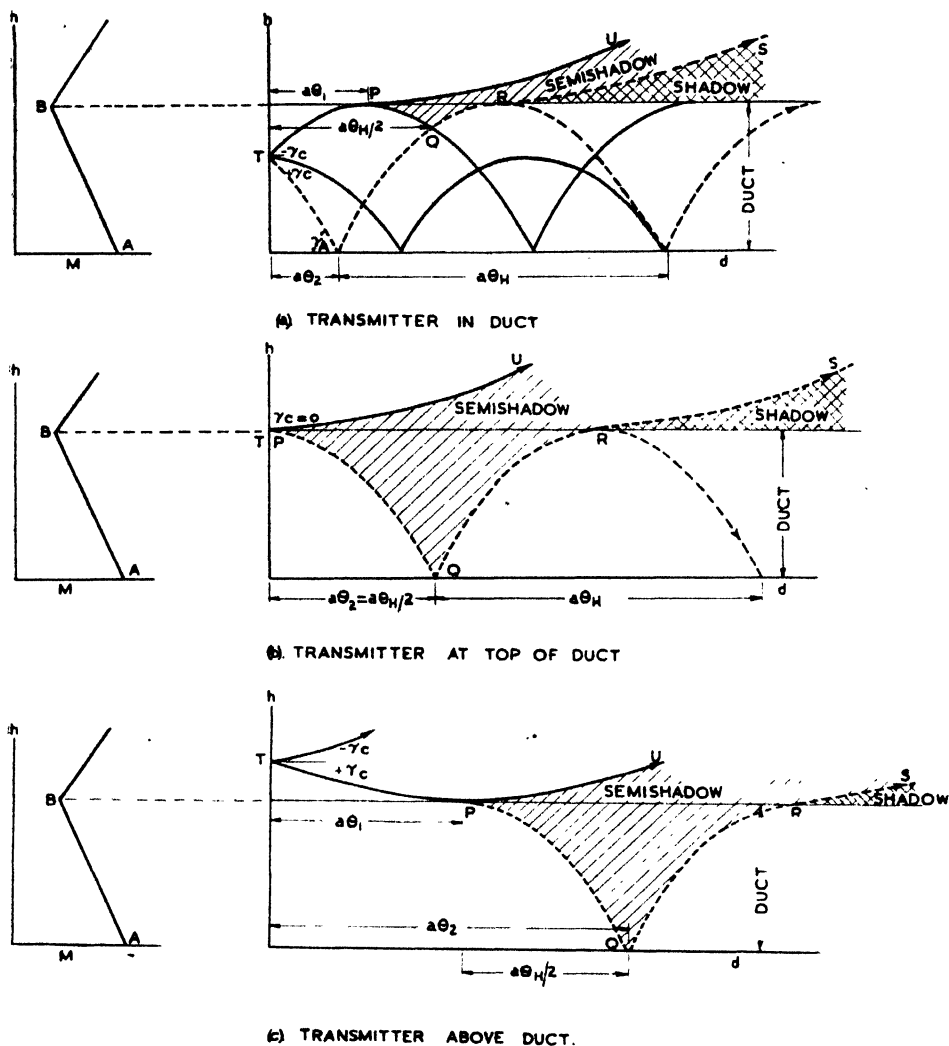


Figure 5. Shadow regions caused by a surface duct.

From a transmitter below B (figure 5 (a)) there will be a pair of rays leaving at angles $\pm \gamma_c$, which just reach the level of the minimum of N , that at $+\gamma_c$ doing so after reflection from the earth. The flux emitted between these rays is trapped between the earth and the horizontal through B, this region being called a surface duct (Booker and Walkinshaw 1946).

The upper critical ray, initial angle $-\gamma_c$, reaches the level of B at a finite distance, whose angular range θ_1 is given approximately by

$$\theta_1 = \int_{-\gamma_c}^0 -k d\gamma / (a+h) \simeq -k\gamma_c/a; \quad (k \text{ constant, } h \ll a), \quad \dots\dots (25)$$

$$\gamma_c = +\sqrt{2(N_T - N_B)^{\frac{1}{2}}}. \quad \dots\dots (26)$$

Then it returns to earth in a half-hop, is reflected, and continues between earth and level B in full hops around the earth. The lower critical ray, $+\gamma_c$, reaches the earth at an angular range θ_2 given by

$$\theta_2 = \int_{\gamma_c}^{\gamma_A} -k d\gamma / (a+h) \simeq -k(\gamma_A - \gamma_c)/a; \quad (k \text{ constant, } h \ll a). \quad \dots\dots(27)$$

$$\gamma_A = +\sqrt{2(N_A - N_B)^{\frac{1}{2}}}. \quad \dots\dots(28)$$

The angular range of the critical ray hop is

$$\theta_H = 2(\theta_1 + \theta_2) \simeq -2k\gamma_A/a; \quad (k \text{ constant, } h \ll a). \quad \dots\dots(29)$$

Each critical ray may be considered to divide at P, R, one branch continuing within the duct as described and the other proceeding into the region of positive k , above B.

The angles of elevation at P, R are zero. Between the two upper branches is a lamina region of "semi-shadow" into which the direct rays do not penetrate, but which is traversed by reflected rays leaving the transmitter at angles greater than γ_c . The tilt of the lamina increases with height, but for a given height of transmitter the angular width of the shadow remains constant at $2\theta_2$.

It is apparent that this width will be zero when the transmitter is at the surface but that the amount of trapped radiation is then a maximum, since γ_c is a maximum. As the transmitter is raised to the top of the duct (B), the trapped radiation decreases to zero, the width of the shadow increases to θ_H but the angular distance to the centre, measured at the level B, remains constant at $\theta_H/2$ (figure 5 (b)).

The lower boundary of the semishadow PQR lies below the duct, the area PQR increasing with the height of the transmitter till the top of the duct is reached.

If the transmitter is raised above the duct, trapping does not occur but the shadow zone is still formed. The lamina angular thickness now remains constant at θ_H , but its centre moves away from the transmitter with increase of altitude (figure 5 (c)).

Beyond the lamina and above the duct is a full shadow zone not reached by direct or reflected radiation.

The above description presents the ideal distribution given by ray theory. As stated in §2.1 the treatment can be made more quantitative to limited ranges determined by the practical consideration of the maximum number of intersecting rays which can be handled by interference methods. At long ranges, within the duct, use has been made of the theory of guided waves, the minimum range to which the field may be calculated being determined in this case by the maximum number of modes which can conveniently be dealt with. Considerable work has been done in this way on surface trapping. Intermediate ranges are not amenable to either method and little work by way of calculation has been done for the semishadow region.

It is clear that some radiation will be diffracted into the "semishadow" and "shadow" across the curved surfaces PU and RS, and that some will penetrate the boundary PQR, in accordance with wave theory, but the diffracted field will generally be weak compared with the radiation which would reach the space in the absence of trapping. Neglecting then the diffracted field in order to obtain an upper limit to the decrease in intensity, it is seen that the normal interference field at the maxima of the lobes will be decreased about 6 db. in the semishadow

region, due to the absence of the direct ray. (This does not take into account increased divergence of the reflected ray due to the negative gradient.) Such a decrease would be important when searching for aircraft at long and medium ranges, since a radar set is then working near the limit of its ordinary detection range.

Ray treatment predicts a second major decrease in field strength on passing into the full shadow across the false horizon RS—theoretically to zero. The problem of calculating the residual field beyond RS is similar to that for the case of diffraction around the earth's surface near the horizon when superrefraction is absent, but is complicated by leakage across the diffracting boundary RV. A full wave theory is required.

The (M, h) curves upon which the discussion has been based are drawn with discontinuities of gradient at the maxima and minima, and with constant gradient between. Actual soundings show some rounding of the turning points and curvature of the joining sections, but as will be seen from figure 8 (b), this is often small. The effect of rounding of the minima is to make $\text{grad } N \rightarrow 0$, and $d \rightarrow \infty$. The critical rays tend asymptotically to approach the horizontal through B, but the ray for which the value of γ at level T is slightly greater than γ_c cuts the top of the duct at a finite range. The region bounded by the two rays at $\pm \gamma$ will only contain a small amount of direct radiation above the duct, proportional to $\gamma - \gamma_c$, and may be regarded as the semishadow zone (cf. figure 8 (c)). One result is that the shadow boundaries will be diffuse instead of sharp, a condition which is aided by diffraction.

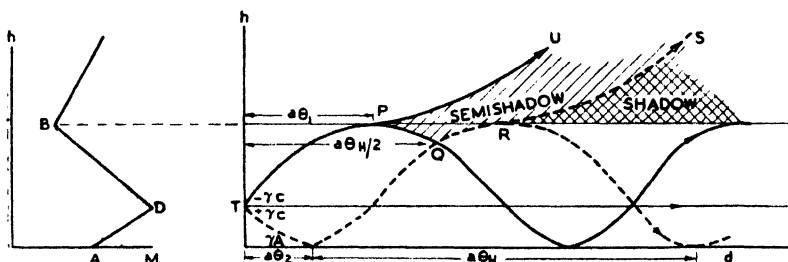


Figure 6. Surface duct. Elevated layer. Transmitter at max. M .

(c) *Figure 6.* The initial gradient is positive and is followed by the duct-forming negative gradient with a minimum N less than the surface value. The rays below D will be concave up, and maximum trapping occurs with the transmitter at D. Conditions otherwise vary much as in (b).

(d) *Figure 7.* As in (c), a positive gradient precedes a negative gradient, but the minimum N now is greater than the surface value. This is an elevated duct.

Here the lower critical ray descends only to the level A, where N has the same value as at B, so that the duct lies between the horizontals A and B. Interesting features are (i) the caustic passing through R, extending down to the level of the transmitter and up into the shadow; and (ii) the extension of semishadow beyond R, the full shadow boundary now being the continuation of the ray tangent to the earth beyond V. The rays whose intersections form the caustic lie between this tangent ray and the ray which leaves the transmitter horizontally.

When the transmitter is below the duct all radiation passes through it, so that there is no trapping or shadow. As the transmitter moves up through the duct, zero trapping occurs at A as well as at B, maximum trapping at D. The sequence of variation of shadow width and centre are similar to (b) and (c).

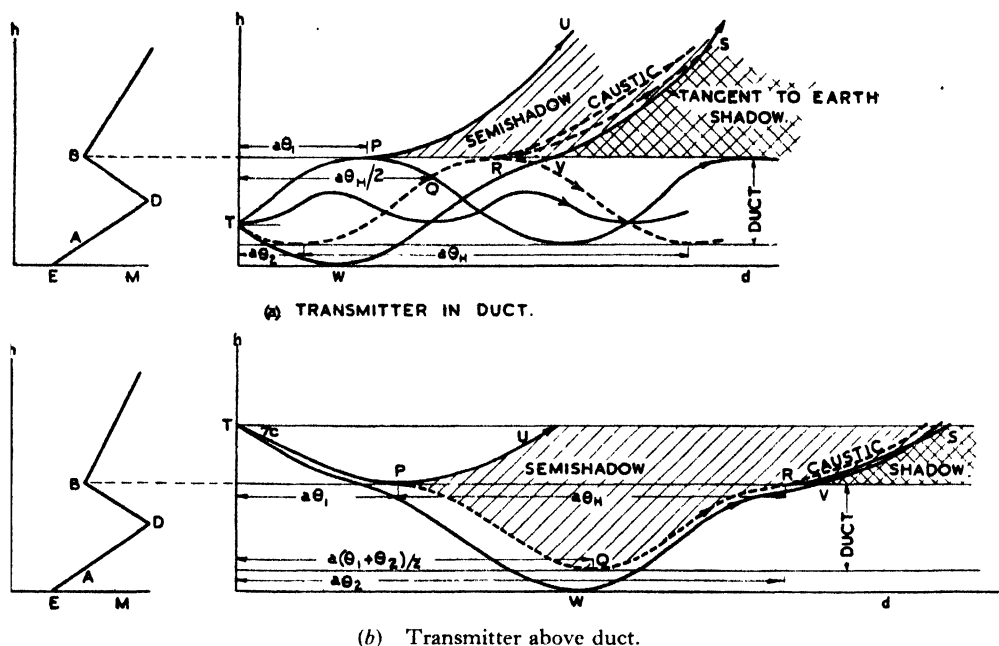


Figure 7. Shadow caused by an elevated duct.

§ 4. INCLINED LAYERS

When the stratification is inclined, but concentric, the condition for possible trapping is that $(N + \int dr'/r')$ must have a negative gradient. We now have

$$\gamma'_c = \sqrt{\{-2\delta(N + \int dr'/r')\}},$$

the increment δ being taken from transmitter level to the top of the duct.

If the position of the centre of curvature is changing, the negative gradient of $(N + \int d_1 n'/r')$ must produce a change in elevation angle $\delta_1 \gamma'$ greater than the rotation $\delta_2 \gamma'$ due to the shift in curvature, in order that trapping may occur. In this case there is the possibility that rays trapped at one section may later penetrate the trapping layer owing to a change in its tilt. Again, rays initially above the minimum of $(N + \int d_1 n'_1/r')$ may be trapped on account of a favourable change in the tilt.

§ 5. EXPERIMENTAL FLIGHTS IN THE SUMMER OF 1946

5.1. Locality and equipment

In January 1946, a series of meteorological soundings were made in combination with radar measurements on a 200-Mc/s. C.O.L. Mark V air warning set at the radar station at Collaroy, 400 ft. above sea level and 15 miles north of the Sydney Heads. The aircraft flew about 100 miles out to sea at right angles to the coast and back, in a sequence of ascents and descents, to heights of 2000 to

3000 feet. On three of these flights observations were made on the radar set of the time, and of the range, bearing and signal-to-noise ratio of the echo from the plane. Observations of range were made every 10 seconds and of bearing every 2–3 minutes. The power output was monitored but an absolute calibration was not attempted.

In the aircraft a simultaneous record was made of time, number of the reading, air speed by air speed indicator, height, and hence pressure from an altimeter, and wet and dry wick temperatures, using thermocouples. The indicating meters were mounted on one panel and photographed at controllable intervals (3 to 6 seconds) by a camera with motor-driven shutter. The thermocouple and camera were designed and built by M. Iliffe and will be described elsewhere (Iliffe 1948). The hygrometer was mounted forward of the cabin and above the fuselage clear of slipstream disturbance and engine contamination. Times were synchronized with the radar station. Air speed and rates of climb were maintained as constant as possible by the pilot.

The flights were made on days forecast as favourable to strong *M*-inversions. At this time of the year the system of eastward moving anticyclones, which form the central motif of the Australian weather pattern, has a general track which crosses the eastern coast about latitude 40° S. and reaches the latitude of Sydney in the central Tasman Sea. The circulation associated with the rear portion of the anticyclone after it crosses the coast, creates strong off-shore streaming of very hot dry air. Modification by the cooler sea from below tends to produce the familiar surface advection duct with strong and deep temperature and humidity inversions. A further diurnal modification is produced on the circulation by the hot land-cool sea influence on the density and pressure gradients. This results in a large-scale sea breeze effect which produces a change in the shape of the temperature and humidity contours. The reversal of wind direction near the sea allows the surface layer of modified air to become cooler and moister, and more thorough mixing is achieved to greater height and closer inshore. The transition layer between modified and unmodified air becomes thinner and higher and an elevated duct is formed. The contours of constant temperature and humidity slope upwards out to sea, with a gradient which increases slightly from the land, producing an upward curvature. A further effect of the reversal of low level circulation is to slow down the eastward movement of the anticyclone, causing the inversion conditions to persist for some days.

The detailed history of the development of these "coastal fronts" (Martyn and Squires 1944) is of considerable interest, and this aspect is to be investigated further by the Radiophysics Laboratory. In this paper, however, the interest is in the illustration afforded of the decrease in field strength across the transition layer between the lower and upper air masses. The elevated radio ducts so formed have *M*-inversion layers of exceptional magnitude and steepness, changes in *M* of 30 units sometimes occurring in 150 feet.

5.2. *Experimental results of the soundings*

5.2.1. *Experimental results* for two flights made on the 10th and 11th January, are presented in figures 8 (a), 8 (b), and figures 9 (a), 9 (b) respectively and the ray analysis for the first flight in figures 8 (c), 8 (d). Figures 8 (a) and 9 (a) give the contours of constant *M* for the outward run, and figures 8 (b) and 9 (b) give (*M*, *h*) curves at each 10 miles of this path. Thus (a) and (b) present the vertical

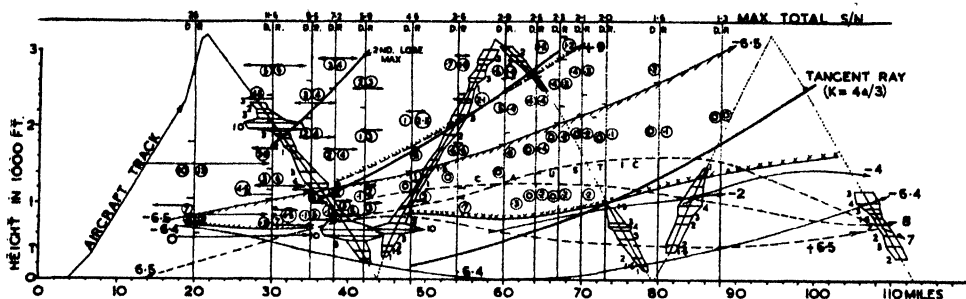


Figure 8 (d). Values of S/N calculated from divergence of rays in figure 8 (c).

Figures in circles are calculated values of average S/N at range of height lines beside which they are written; average calculated for the height between the short horizontal lines. These values are represented by the horizontal lines terminated by arrow heads to the same scale as the measured S/N ; height lines are dotted when total $S/N < 1$. Figures at top of height lines are twice the calculated free space S/N for the given range.

Top of semishadow |||||
Top of shadow |||||

Bottom of semishadow |||||
Bottom of shadow x x x x

D = direct, R = reflected.

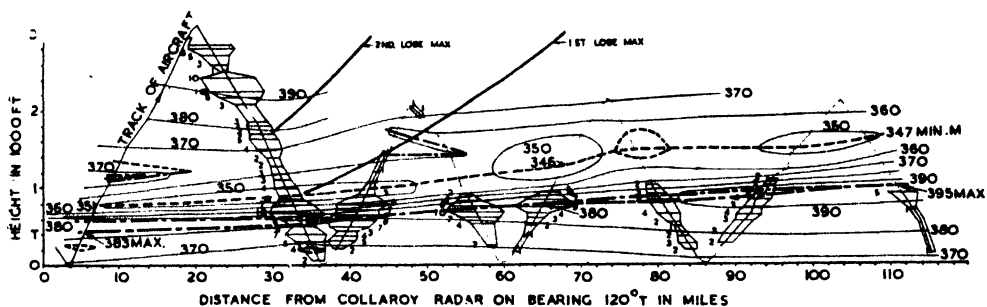


Figure 9 (a). Contours of $M = \text{constant}$ for aircraft sounding, outward flight, 11 Jan. 1946, from Collaroy, Sydney, N.S.W. For details see figure 8 (a).

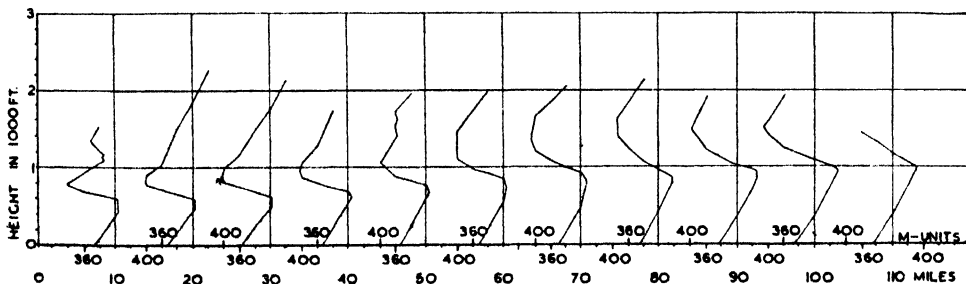


Figure 9 (b). M, h curves at a number of ranges from the stratification of figure 9 (a).

section through the atmosphere, at right angles to the coast, which coincides with the axial plane of the transmitter.

The track of the aircraft is shown in figures 8 (a) and 9 (a). The variation of the signal-to-noise ratio (S/N) is indicated graphically by drawing two lines, one on each side of the track, at a distance apart proportional to S/N . Values of S/N are written alongside. Where no echo was detected the track is shown dotted. For comparison, the positions of the first and second theoretical lines of maximum field for $k=4a/3$ are also drawn.

The ray diagram 8 (c) was computed from figures 8 (a) and 8 (b) and from this were calculated the theoretical values of S/N shown in figure 8 (d).

5.2.2. *Accuracy of the data.* The values of $(\mu-1)$ were calculated from Stickland's (1942) formula:—

$$\mu-1 = (79P/T - 11e/T + 3.8 \times 10^5 e/T^2) 10^{-6} \quad \dots\dots (30)$$

the constants being rounded off to these values. P is the total pressure of the air, e the partial pressure of the water vapour (pressures in millibars) and $T^\circ \text{K}$. the absolute temperature of the air.

Temperatures were read to $\pm 0.1^\circ \text{C}$. and corrected for plane speed. Apart from the question of the accuracy of the formula constants, it is considered that the overall accuracy in M is $\pm 1 M$ -unit, so far as individual readings are concerned. Accuracy of the contours is affected by the time factor involved in the duration of a flight and by the accuracy of the determination of position. The outward journey took about 50 minutes. A comparison made with the return flight (not given here) showed that the change in structure during the time of flights was small enough to justify using the contours for field strength calculations, so far as the time factor is concerned. Drawing of the contour lines by joining points, in some cases 20 miles apart, could not allow for small scale undulations and undoubtedly the curves could not be as smooth as shown. However, the fact that three consecutive points generally lie near a straight line would indicate that large scale undulations did not occur. Curves (b) were plotted from the contours of (a), and so again are somewhat smoothed. Bearing on this point, which involves the scale of the temporal and spatial homogeneity of the atmosphere is the fact that the original (M, h) curves taken along the slant track of the plane were, in fact, comparatively smooth. At the lower levels readings were usually made at height intervals of 15–20 feet and departure from the degree of smoothing represented by (b) seldom exceeded the accuracy of the M values, $\pm 1 M$ -units. A point of interest in the curves (b) is the sharpness of the turning points where the gradients of M change sign.

5.3. *Analyses of the soundings*

5.3.1. *Comparison of S/N with contours of M constant* (figures 8 (a), 9 (a)). Examination of the figures reveals the following broad features also possessed by the other flights of the series:—

(a) The field strength, as indicated by the signal-to-noise ratio has a value which remains high from a point about the middle of the first inversion down to near sea-level, where the aircraft becomes lost to view. A strong maximum occurs near the first maximum of M (usually a little below it) so that the position of maximum field lies along a line which follows the inversion upwards as the latter rises out to sea.

(b) The field falls off rapidly shortly after the inversion is entered from below, sometimes to a value below receiver threshold, through a height interval which increases with range; above this S/N increases with height. This is shown by the values of S/N obtained during the second ascent and the first and second descents of the plane. The later climbs were not high enough to show the final increase of field. The upper boundary of the region of diminished field is thus an upward sloping line whose location is shown in § 5.3.2 to be near the upper critical rays. The lower boundary is within the M inversion.

(c) The first lobe has disappeared in figure 9 (a). In figure 8 (a) there appears what could be an interference maximum representing the first lobe displaced upwards about 150 feet from its position with $k=4a/3$, and another at 57 miles 2650 feet, this time displaced upwards 600 feet. A neighbouring maximum at 55 miles would appear to be due to passage across the shadow boundary at this point.

(d) Above the upper boundary the field tends to resume its normal lobe structure as indicated by maxima at 30 miles 2050 feet in figure 8 (a) and 29 miles 1650 feet in figure 9 (a), representing the second lobe. In figure 8 (a) the position of the second lobe, like that of the first lobe, is high, while in figure 9 (a) the second lobe appears in its normal position.

5.3.2. *Comparison of S/N with ray tracing.* Rays were traced for the first sounding, covering the range of angles of elevation measured at the transmitter from 10×10^{-3} radians to -10×10^{-3} radians. These are plotted in figure 8 (c). The direct rays (negative α) are shown as full lines and the reflected rays (positive α) as dotted lines. The rays will be referred to by their initial angles of elevation in units of 10^{-3} radians. The lines of maximum and minimum M and the variation of S/N along the aircraft track are repeated from figure 8 (a).

It will be noticed that the elevated duct of figures 8 (a), 8 (b), is double, to a range of about 75 miles. The first M inversion, which falls to a deep minimum, is the dominant factor affecting the ray paths. This inversion is followed by a smaller duct whose influence is mainly noticed on the reflected rays. There is a further complication in the M -structure between 40 and 90 miles commencing at 2000 feet where another M -inversion occurs. On account of its height this has only a minor effect.

Both the direct and the reflected rays are strongly trapped by the first inversion. Splitting of the direct radiation occurs at about 11 miles 700 feet, the -6.5 ray penetrating all the inversions while the -6.4 ray is trapped by the first inversion (see figure 8 (d)).

The critical rays are not so clearly defined for the reflected radiation. The first inversion separates the $+6.5$ ray from the $+7$ ray at about 40 miles 850 feet, and the critical ray would be about $+6.8$. However, rays up to $+8$ become trapped by the second inversion, trapping being aided by the rise in the position of the inversion. This results in the full shadow being divided into two regions of diminished field, separated by a caustic. The upper region extends to about the $+9$ ray, but overlaps the semishadow, the two top edges crossing at 36 miles.

An estimate was made of the limits of the field strength values above the duct in terms of S/N , from a consideration of the divergence of the rays. It is felt that the experimental figures obtained from the radar measurements did not warrant closer calculations by taking into account interference effects or by

using the extension of ray theory mentioned in §2.1, but at the same time the results correlate well enough to suggest that more detailed field strength measurements would justify fuller analyses. The expected S/N ratio for the direct and the reflected rays was estimated from the approximate relations:

$$S_D = \frac{S_0 \alpha_D}{h_D d_D}; \quad S_R = \frac{S_0 \alpha_R}{h_R d_R}, \quad \dots\dots (31)$$

$$S_{\max} = S_D + S_R; \quad S_{\min} = |S_D - S_R|, \quad \dots\dots (32)$$

where S = average S/N ratio at range d , between two rays initially inclined to each other at α ,

h = height interval between the two rays at range d and initial inclination α ,

S_0 = value of S in free space at unit range.

The subscripts D and R stand for the direct and the reflected rays respectively. In equation (31) the earth's divergence factor is neglected and it is assumed that its reflection coefficient is unity. The value of S_0 was obtained by assuming that S_D , S_R are in phase at 30 miles 2000 feet, where the measured signal-to-noise ratio passes through a maximum, with $S_D + S_R = 10$.

The results are shown in figure 8 (*d*). In this figure vertical lines are drawn at a number of ranges and on these lines are marked their points of intersection with the rays. The average values of S_D , S_R , calculated for the height intervals between the intersection points, are written in circles each in its proper interval alongside the height line, and actual values of S_D and S_R are represented by horizontal lines terminated by arrowheads placed against their written values and drawn to the same scale as the representation of the measured values of S/N . Where $S_D + S_R < 1$ the height line is shown dotted. In this way the two quantities, calculated maximum $S/N (= S_D + S_R)$ and measured S/N may be visually compared. The number written at the top of each height line is twice the free space value for S/N at the particular range. The critical rays mentioned in this section are shown by shading. The field at the caustic formed within the shadow region cannot be determined by this method of calculation.

The lower boundary of the shadow, which is also the upper boundary of the radio duct, has been placed some 50–100 feet above the envelope of the rays trapped by the first inversion.

Comparison of the measured and calculated fields shows that

- (a) The position of the calculated shadow boundaries agrees quite well with the radar results;
- (b) The measured values of S/N above the duct are, in general, in good accord with the requirement that they lie between $S_D + S_R$ and $|S_D - S_R|$.

Discrepancies which occur at 38 miles and to a lesser extent at 48 miles within the semishadow are larger than the experimental limits and deserve examination. The recorded values of the minimum of S/N occurring at each of these places is 1, whereas the calculated values of $(S_D - S_R)$, although passing through a minimum, only fall to 4 in the first case and 2 in the second. The field is almost entirely due to the reflected ray. A possible explanation of this discrepancy is that it occurs at the commencement of the semishadow where it is narrow, and that edge effects which tend to make the boundaries diffuse may extend to a big proportion of the total thickness. Thus leakage from the

direct rays across both boundaries and partial reflection of the reflected rays at the strong M inversion would both tend to decrease the value of S/N below the lower calculated limit $|S_D - S_R|$.

The methods used cannot of themselves give a quantitative estimate of the complex field distribution within the duct. However, the experimental fact that the signal is a maximum near the maximum of M is in accord with the bunching of rays and the occurrence of caustics at several places near these maxima. A full delineation of the caustics and their cusps would require the drawing of considerably more rays; the intersection of rays belonging to the same pencil may be observed at the following places, points in figure 8 (c): direct rays 20 to 35 miles, about 400 feet; 60 to 85 miles, 800–1100 feet; reflected ray, 55 miles (intersection of ray $+6$ and $+6.5$).

Although a quantitative estimate of the field distribution within the duct is not feasible by these methods a calculation can be made of the average S/N across the duct at a given range on the assumption that the energy remains trapped without appreciable leakage, but is attenuated through horizontal spreading. It is of interest to note that an average value calculated in this way does lie within the limits observed for S/N at the same range. For instance assuming the angle of the trapped radiation to be 13×10^{-3} radians (-6.5 to $+6.5$) and the vertical spread at 110 miles to be 1300 feet (say from 300 feet to 1600 feet) the theoretical value of S is 2.5; the measured values vary from 4 to 1.5.

§ 6. CONCLUSIONS

The results of the last section appear to confirm, for frequencies around 200 Mc/s., earlier theories of the existence of shadow zones during certain predictable weather conditions. The results also indicate the utility and practicability of ray tracing methods for obtaining the broad outlines of the field distribution.

It is suggested that those using high-frequency radio systems for the detection or guidance of aircraft in localities subject to strong inversions should keep in mind the possibility, at times, of failure in the system from this cause.

ACKNOWLEDGMENTS

Some of the work described in this paper was carried out as part of the research programme of the Division of Radiophysics, Council for Scientific and Industrial Research, Commonwealth of Australia.

The author wishes to express his great appreciation of the invaluable advice and assistance given by Dr. D. F. Martyn of C.S.I.R. who originated the work. He would like to thank his colleague, Mr. M. Iliffe of the Physics Department, University of Adelaide, who organized the later experimental work, for the many discussions and constructive criticisms which have helped to clarify the problem; and Mr. D. R. Brown, author of O.R.G. (Aust.) Report Number 17, some of whose results are discussed in this paper.

The writer also wishes to thank Dr. E. G. Bowen and Dr. J. L. Pawsey of the Radiophysics Laboratory for making available to him the results of the aircraft soundings and the Laboratory reference and computing facilities.

REFERENCES

- BOOKER, H. G., and WALKINSHAW, W., 1946, *Meteorological Factors in Radio-Wave Propagation* (London: Physical Society), p. 80.
BROWN, D. R., 1943, *Report No. 17, O.R.G. (Aust.)*.

- ENGLUND, C. R., CRAWFORD, A. B., and MUMFORD, W. W., 1938, *Bell Syst. Tech. J.*, **17**, 489.
- FREEHAFFER, J. E., 1943, *Report No. 447, M.I.T. Radiation Laboratories.*
- ILIFFE, M. I. G., 1948, in preparation.
- MARTYN, D. F., 1943, *Report No. 19, O.R.G. (Aust.).*
- MARTYN, D. F., and SQUIRES, P., 1944, *Some Factors causing "Superrefraction" on Ultra-High Frequencies in South West Pacific, A.R.P.C. Bulletin.*
- PRICE, W. L., 1943, *Report No. 20, O.R.G. (Aust.).*
- SCHELLENG, J. C., BURROWS, C. R., and FERRELL, E. B., 1933, *Proc. Inst. Radio Engrs.*, **21**, 427.
- STICKLAND, A. C., 1942, *Dielectric Constant of Water Vapour and its Effect upon the Propagation of very Short Waves*, National Physical Laboratory, Radio Department Report RRB/S2; 1946, *Meteorological Factors in Radio-Wave Propagation* (London: Physical Society), p. 253.

Equivalent Path and Absorption for Oblique Incidence on a Curved Ionospheric Region

By J. C. JAEGER
University of Tasmania

MS. received 31 July 1947

ABSTRACT. The problem of oblique incidence on a curved ionosphere with a Chapman distribution of ionization density is discussed. It is shown that absorption, equivalent path, and the projection of the path on the earth's surface may all be expressed by equivalence theorems of the plane earth type with suitably modified parameters and functions. Numerical values of the functions involved are given.

§ 1. INTRODUCTION

IN a recent paper (Jaeger 1947, referred to below as V) the equivalent path and absorption * for a wireless wave incident vertically on a Chapman region were calculated. This work is extended here to cover oblique incidence on a curved Chapman region in which the ion-density is a function of height only, and the effects of the earth's magnetic field are neglected. Other studies of the oblique incidence problem have either assumed a parabolic distribution of ion-density (Appleton and Beynon 1940, 1947; for other references on the subject see the bibliography in these papers) or have assumed special properties of the region, for example, that it be thin.

The present work consists essentially of an exact solution of the problem of a ray incident obliquely on a curved Chapman region. Since, even for vertical incidence, a final stage of numerical integration is necessary, the accurate integrals for oblique incidence in any particular case are not much more difficult to evaluate than any approximations to them. An approximation is introduced below in order to simplify the presentation of the work by reducing the number of parameters in terms of which the solution is expressed. At all stages the validity of

* Absorption in a Chapman region was first studied by Appleton (1937) who discussed the case of penetration of a region in which μ is effectively unity, and obtained the formula V (26) which I attributed to Best and Ratcliffe (1938)—not (1928) as in V. He also considered the case of V § 8 in which μ^2 may be comparable with f^2 , and gave the result V (45) with a graph of the integral involved but not the explicit expression V (48) for it.

the approximation has been checked by evaluating the accurate integrals and the discrepancy found to be negligible.

It is found that critical frequency, absorption, equivalent path, and distance traversed along the earth's surface can all be expressed by simple generalizations of the equivalence theorems deduced by Martyn (1935) for a plane ionosphere.

§ 2. FUNDAMENTAL FORMULAE

We have to consider the path of a ray in an ionized region whose ion-density is a function only of h , the height of the point considered above the earth's surface. Let i be the angle of incidence of the ray at height h , and α that at zero height, that is, at the earth's surface (figure 1). Then if R is the radius of the earth, and μ the refractive index at height h , the fundamental equation for refraction in a curved stratified region is

$$\mu(R+h) \sin i = \text{constant} = R \sin \alpha. \quad \dots\dots(1)$$

All the integrals which will be required below involve the quantity

$$\frac{ds}{\mu} = \frac{dh \sec i}{\mu} = \frac{dh}{\sqrt{\{\mu^2 - [R/(R+h)]^2 \sin^2 \alpha\}}}, \quad \dots\dots(2)$$

where ds is an element of the path of the ray. We proceed to study the quantity (2) in detail; when this has been done the evaluation of the integrals is an easy matter.

The refractive index μ is given by the relation

$$\mu^2 = 1 - N\epsilon^2/\pi m f^2, \quad \dots\dots(3)$$

where f is the frequency of the wave, ϵ and m are the charge and mass of the ions, and N is the ion-density.

In a Chapman region N is given by

$$N = N_0 \exp \frac{1}{2}(1 - z - \sec \chi e^{-z}), \quad \dots\dots(4)$$

where χ is the zenith angle of the sun (assumed constant over the region of the ionosphere concerned), and

$$z = (h - h_0)/H \quad \dots\dots(5)$$

is the height above the datum level h_0 measured in terms of the scale height H of the region (cf. V § 2, or Chapman 1931). The quantities N_0 , H , and h_0 may be obtained from analysis of vertical incidence data by a method outlined in V. N_0 is the maximum possible value of N , and occurs at height h_c at the equator at noon at the Equinox. The critical frequency at vertical incidence, f_c , was found in V (9) to be

$$f_c^2 = (N_0 \epsilon^2 / \pi m) \cos^2 \chi. \quad \dots\dots(6)$$

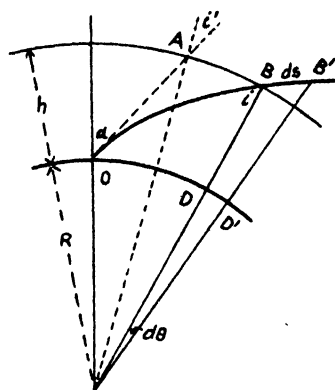


Figure 1.

Using (3), (4), and (5) in (2), and making the substitution

$$y = (\frac{1}{2} \sec \chi)^{\frac{1}{2}} e^{-\frac{1}{2}z}, \quad \dots\dots(7)$$

so that

$$z = \ln(\frac{1}{2} \sec \chi) - 2 \ln y, \quad \dots\dots(8)$$

(2) becomes

$$\frac{ds}{\mu} = -\frac{2H}{y} \frac{dy}{y} \left\{ 1 - by \exp(-y^2) - \frac{R^2 \sin^2 \alpha}{[R + h_0 + H \ln(\frac{1}{2} \sec \chi) - 2H \ln y]^2} \right\}^{-\frac{1}{2}} \quad \dots\dots(9)$$

where, as in V (14), we write

$$b = (f_c^2/f^2) \sqrt{(2e)}. \quad \dots\dots(10)$$

To reduce (9) to a more manageable form, we introduce new parameters β , I , d , and a , defined by

$$\beta = 1 + h_0/R + (H/R) \ln(\frac{1}{2} \sec \chi), \quad \dots\dots(11)$$

$$\sin I = (\sin \alpha)/\beta, \quad \dots\dots(12)$$

$$d = b \sec^2 I = (f_c^2/f^2) (2e)^{\frac{1}{2}} \sec^2 I, \quad \dots\dots(13)$$

$$a = (4H/\beta R) \tan^2 I. \quad \dots\dots(14)$$

In terms of these, (9) becomes

$$\frac{ds}{\mu} = -\frac{2H}{y} \frac{dy \sec I}{y \Delta^{\frac{1}{2}}}, \quad \dots\dots(15)$$

where

$$\Delta = 1 - dy \exp(-y^2) + \tan^2 I \{1 - [1 - (2H/\beta R) \ln y]^{-2}\}. \quad \dots\dots(16)$$

The expression (15) is still exact; to simplify it further, the orders of magnitude of the parameters involved must be considered. Taking h_0 to be of the order of 120 and 300 km. for E and F regions, respectively, and H to be of the order of 15 and 60 km. in these regions, we find that as χ runs from 0° to 80° , β runs from 1.017 to 1.021 for E region, and from 1.040 to 1.057 for F region. If we allow values of α between 0° and 90° , the maximum values of I given by (12) will be 79.5° and 74° for E and F regions respectively; also I increases steadily to these values as α increases from 0° to 90° .

If $(2H/\beta R) \ln y < 1$, we may use the binomial theorem in (16) and obtain

$$\Delta = 1 - dy \exp(-y^2) - a \ln y \{1 + (3H/\beta R) \ln y + \dots\}, \quad \dots\dots(17)$$

where a has been defined in (14).

Now in all cases $(H/\beta R) < 0.01$, so we may neglect all terms of the series in (17) after the first if $0.03a (\ln y)^2$ is negligible to our order of approximation. This is the case in all the calculations made here, since in the integrals which have to be evaluated there are additional factors (such as $y \exp(-y^2)$) which are small for the large and small values of y for which $\ln y$ is large and the approximation above might be inadequate.

Thus for purposes of calculation we may replace (17) by

$$\Delta = 1 - dy \exp(-y^2) - a \ln y. \quad \dots\dots(18)$$

This has the advantage of expressing the whole of the phenomena in terms of the parameters d and a . a is zero for normal incidence or for a plane ionosphere.

it increases from 0 to about 0.27 for E region, and from 0 to about 0.44 for F region, as α increases from 0° to 90° . We shall study the values 0.2 and 0.4 of a which may be regarded as fairly extreme values suited to E and F regions. In § 3 the parameter d will be related to the critical frequency f_x for oblique incidence at α , and the results can thus be expressed in terms of the two parameters f/f_x and a .

§ 3. THE CRITICAL FREQUENCY

For a ray which is transmitted through the ionosphere, $\mu \cos i$ in (2), and thus Δ in (15), never falls to zero. The critical frequency f_x for a ray leaving the earth's surface at α is the greatest value of the frequency f for which Δ vanishes. Thus for a given value of α , and therefore of a , the critical frequency is to be determined from the least value d_a of d for which Δ vanishes. Using the form (18) of Δ , the condition for this is that the equation in y

$$1 - a \ln y - y d_a \exp(-y^2) = 0, \quad \dots\dots (19)$$

have a double root. This requires that in addition to (19) its derivative shall vanish, that is,

$$a/y + d_a(1 - 2y^2) \exp(-y^2) = 0. \quad \dots\dots (20)$$

From (19) and (20) we find

$$a = \frac{2y^2 - 1}{(2y^2 - 1) \ln y + 1}, \quad \dots\dots (21)$$

$$d_a = [y \exp(-y^2) \{(2y^2 - 1) \ln y + 1\}]^{-1}. \quad \dots\dots (22)$$

These are a pair of parametric equations for a and d_a in terms of the parameter y .

If $a = 0$, $d_a = \sqrt{(2e)}$, and $y = 1/\sqrt{2}$, the values for vertical incidence. For oblique incidence only a small range of values of y greater than $1/\sqrt{2}$ is needed to cover the required range $0 < a < 0.5$ of a , and thus the replacing of (16) by (18) is justified. For the relevant range of β and a the deviation of results calculated from (21) and (22) from those from (16) is very small. For example in the rather extreme case $a = 0.4$, the value of d_a calculated from (21) and (22) is 2.578, while the accurate formula derived from (16) with $a = 0.4$, and $(4H/\beta R) = 0.03628$, gives 2.576.

Values of $[\sqrt{(2e)/d_a}]^\dagger$, calculated from (21) and (22), are shown in figure 2. The critical frequency f_x now follows from d_a by using (13); it is

$$f_x = f_c [\sqrt{(2e)/d_a}]^\dagger \sec I. \quad \dots\dots (23)$$

For a plane earth this becomes $f_x = f_c \sec \alpha$, the plane earth equivalence theorem; this is seen to generalize immediately to the curved earth on replacing α by I and multiplying by the factor $\sqrt{[\sqrt{(2e)/d_a}]}$ of figure 2 which is nearly unity.

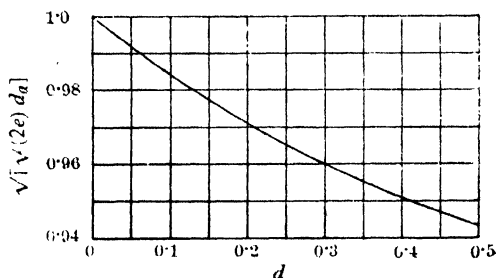


Figure 2.

§4. ABSORPTION IN A DOUBLE PASSAGE OF THE LAYER AT FREQUENCIES GREATER THAN THE CRITICAL FREQUENCY f_x

The total absorption is

$$2 \int \kappa ds, \quad \dots\dots(24)$$

taken through the layer, where, as in V,

$$\kappa = \frac{\nu}{2c} \frac{1-\mu^2}{\mu} = \frac{N_0 \epsilon^2 \nu}{2\pi c \mu m f^2} \exp \frac{1}{2}(1-z-\sec \chi e^{-z}), \quad \dots\dots(25)$$

$$\nu = \nu_0 e^{-z}, \quad \dots\dots(26)$$

and c is the velocity of light. It is assumed that $(2\pi f)^2 \gg \nu^2$.

Using (25) and (26) in (24), with the value (15) and (18) of ds/μ , we find

$$2 \int \kappa ds = H\nu_0 c^{-1} \cos \chi \cos I F(f_x/f; a), \quad \dots\dots(27)$$

where

$$F(f_x/f; a) = 4d \int_0^\infty \frac{y^2 \exp(-y^2) dy}{\sqrt{(1-dy \exp(-y^2)-a \ln y)}}. \quad \dots\dots(28)$$

Alternatively, writing

$$\Phi(f_x/f; a) = d^{-1} \sqrt{(2e)} F(f_x/f; a), \quad \dots\dots(29)$$

(27) may be expressed in the form

$$\begin{aligned} 2 \int \kappa ds &= H\nu_0 c^{-1} \cos \chi \sec I (f_c/f)^2 \Phi(f_x/f; a) \\ &= H\nu_0 c^{-1} (d_a/\sqrt{(2e)}) \cos \chi \cos I (f_x/f)^2 \Phi(f_x/f; a). \end{aligned} \quad \dots\dots(30)$$

For vertical incidence $a=0$, and $F(f_x/f; 0)$ and $\Phi(f_x/f; 0)$ are the functions $F(f_c/f)$ and $\Phi(f_c/f)$ defined in (23) and (25) of V and tabulated there.

For a plane ionosphere $a=0$, $I=\alpha$, and (27) becomes

$$2 \int \kappa ds = H\nu_0 c^{-1} \cos \chi \cos \alpha F(f_x/f; 0). \quad \dots\dots(31)$$

This is just $\cos \alpha$ times the absorption for vertical incidence at the frequency $f \cos \alpha$ as required by Martyn's equivalence theorem (1935). The modification of this result for a curved ionosphere consists of replacing α by I , and the function $F(f_x/f; 0)$ by $F(f_x/f; a)$.

The function $F(f_x/f; a)$ is plotted in figure 3 for the values 0, 0.2, and 0.4 of a , corresponding to vertical incidence (or plane ionosphere) and rather large values for E and F regions respectively.

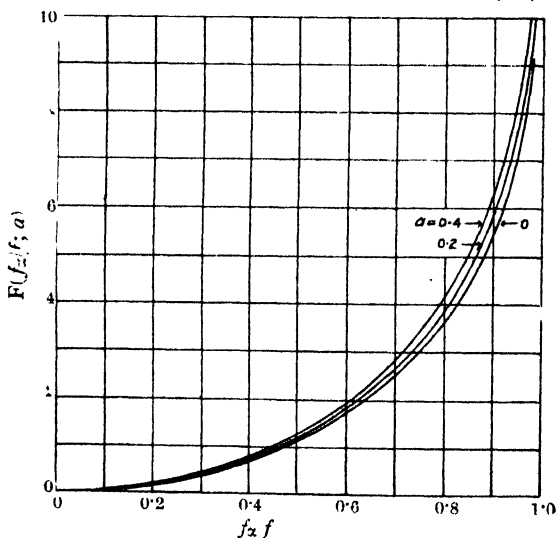


Figure 3. The absorption function for penetration.

§ 5. ABSORPTION FOR A RAY REFLECTED FROM THE LOWER SIDE OF THE LAYER

In this case we require $2 \int \kappa ds$ over a range from ground level to the highest point of the path. This point is determined by the value z_1 of z corresponding to the greater root y_1 of

$$1 - dy \exp(-y^2) - a \ln y = 0. \quad \dots\dots(32)$$

The values of y_1 for the values 0, 0.2, and 0.4 of a are shown in the dotted curves of figure 6. Using (5) and (8), the height h_1 of the highest point of the path above the earth's surface is found to be

$$h_1 = h_0 + H \ln(\frac{1}{2} \sec \chi) - 2H \ln y_1. \quad \dots\dots(33)$$

Proceeding as in § 4, we find for the absorption

$$2 \int \kappa ds = H \nu_0 c^{-1} \cos \chi \cos I F_1(f/f_x; a), \quad \dots\dots(34)$$

where

$$F_1(f/f_x; a) = 4d \int_{y_1}^{\infty} \frac{y^2 \exp(-y^2) dy}{\sqrt{(1 - dy \exp(-y^2) - a \ln y)}}. \quad \dots\dots(35)$$

In figure 4 the function $F_1(f/f_x; a)$ is plotted for the values 0, 0.2, and 0.4 of a . Small values of f are not included, since for these the simple approximation (18) begins to break down. There is no difficulty in extending the curves to small values of f , but there is not much point in doing so because we have also assumed that $(2\pi f)^2 \gg \nu^2$.

It should be remarked that the fact that the curves of figure 4 cross and are rather more widely separated than those of the other figures is caused by the multiplying factor d in the definition (35) of $F_1(f/f_x; a)$. Curves of $[\sqrt{(2e)/d}] F_1(f/f_x; a)$, analogous to Φ of (29), do not cross and lie close together.

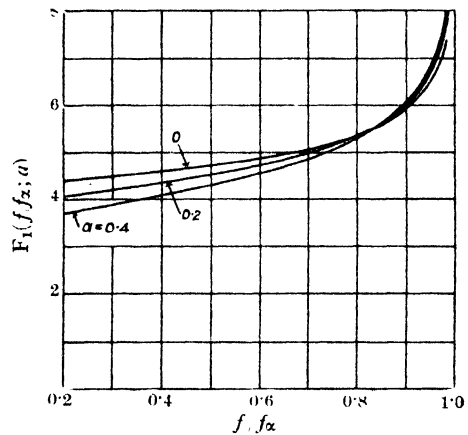


Figure 4. The absorption function for reflection.

§ 6. THE EQUIVALENT PATH

This is given by

$$\int \frac{ds}{\mu} \quad \dots\dots(36)$$

taken over the whole trajectory of the ray. In V the procedure of evaluating the excess of equivalent path over distance traversed was adopted, and in this way a convergent integral was obtained which was easy to evaluate. A similar simplification can be made for oblique incidence by calculating the excess of the equivalent path over the distance measured to the same height along the tangent to the path at ground level. This distance can be calculated by simple trigonometry.

In figure 1 let OB be the path and OA the tangent to it at ground level. Let i' be the angle this tangent makes with the radius vector at height h , so that

$$(R + h) \sin i' = R \sin \alpha. \quad \dots\dots(37)$$

The excess of equivalent path over distance measured along OA up to any height h is now

$$\int_0^h \left\{ \frac{dh \sec i}{\mu} - dh \sec i' \right\}. \quad \dots\dots(38)$$

Using (15) and (18), and (37) with the approximation leading to (18), this becomes

$$2H \sec I \int_y^\infty \left\{ \frac{1}{(1-a \ln y - dy \exp(-y^2))^{\frac{1}{2}}} - \frac{1}{(1-a \ln y)^{\frac{1}{2}}} \right\} \frac{dy}{y}. \quad \dots\dots(39)$$

For a ray transmitted through the region (which by (1) and (37) emerges with $i=i'$) the excess of equivalent path over distance measured along OA for a double traverse of the region is

$$H \sec I P(f_\alpha/f; a), \quad \dots\dots(40)$$

where

$$P(f_\alpha/f; a) = 4 \int_0^\infty \left[\frac{1}{(1-a \ln y - dy \exp(-y^2))^{\frac{1}{2}}} - \frac{1}{(1-a \ln y)^{\frac{1}{2}}} \right] \frac{dy}{y}. \quad \dots\dots(41)$$

Curves of $P(f_\alpha/f; a)$ for $a=0, 0.2, 0.4$ are shown in figure 5; that for $a=0$ is $P(f_c/f)$ of V, so it appears that the effect of oblique incidence is to introduce the factor $\sec I$ in (40) and to change from $P(f_\alpha/f; 0)$ to $P(f_\alpha/f; a)$.

If the wave is reflected from the region, the excess of equivalent path in the whole trajectory over twice the distance measured along OA to the level of reflection is

$$H \sec I P_1(f/f_\alpha; a), \quad \dots\dots(42)$$

where

$$P_1(f/f_\alpha; a) = 4 \int_{y_1}^\infty \left\{ \frac{1}{(1-a \ln y - dy \exp(-y^2))^{\frac{1}{2}}} - \frac{1}{(1-a \ln y)^{\frac{1}{2}}} \right\} \frac{dy}{y}, \quad \dots\dots(43)$$

and y_1 is given by (32). Values of $P_1(f/f_\alpha; a)$ and y_1 are shown in figure 6.

The distance along OA to the level of reflection follows immediately from the trigonometry of figure 1 using the value of h_1 given by (33).

§ 7. THE PROJECTION OF THE PATH OF THE RAY ON THE EARTH'S SURFACE

From figure 1 it follows that the projection DD' on the earth's surface of the element of path BB' = ds of the ray is

$$\frac{R ds \sin i}{R+h} = \frac{R^2 dh \sin \alpha \sec i}{\mu (R+h)^2}, \quad \dots\dots(44)$$

using (1).

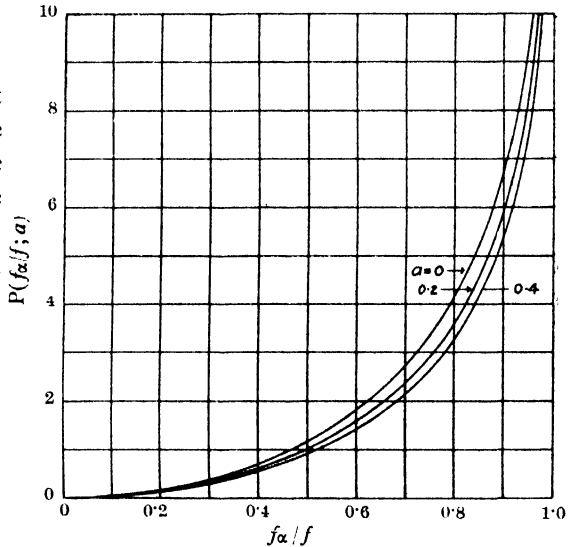


Figure 5. Excess equivalent path for penetration.

Also, the projection on the earth's surface of the element of OA between the heights h and $h+dh$ is

$$\frac{R dh \tan i'}{R+h} = \frac{R^2 dh \sin \alpha \sec i'}{(R+h)^2} \quad \dots\dots (45)$$

Thus the horizontal projection on the earth's surface of the ray up to height h is greater than the horizontal projection of the portion of the tangent to the ray at ground level up to height h by the amount

$$\sin \alpha \int_0^h \left\{ \frac{\sec i}{\mu} - \sec i' \right\} \frac{R^2 dh}{(R+h)^2} \quad \dots\dots (46)$$

As in (39), this is equal to

$$\frac{2H}{\beta^2} \sec I \sin \alpha$$

$$\times \int_y^\infty \left\{ \frac{1}{(1-a \ln y - dy \exp(-y^2))^{\frac{1}{2}}} - \frac{1}{(1-a \ln y)^{\frac{1}{2}}} \right\} \frac{dy}{y[1-(2H/\beta R) \ln y]^2} \quad \dots\dots (47)$$

Thus for a ray transmitted twice through the region, the projection of the path on the earth's surface is greater than the projection of the portion of tangent to the ray at ground level by the amount

$$(H/\beta^2) \sec I \sin \alpha Q(f_\alpha/f; a; 2H/\beta R), \quad \dots\dots (48)$$

where

$$Q\left(\frac{f_\alpha}{f}; a; \frac{2H}{\beta R}\right) = 4 \int_0^\infty \left\{ \frac{1}{(1-a \ln y - dy \exp(-y^2))^{\frac{1}{2}}} - \frac{1}{(1-a \ln y)^{\frac{1}{2}}} \right\} \frac{dy}{y[1-(2H/\beta R) \ln y]^2} \quad \dots\dots (49)$$

The integral Q defined in (49) differs only from P defined in (41) by the factor $[1-(2H/\beta R) \ln y]^{-2}$ in the integrand. Since, for transmission through E region, $2H/\beta R$ is of the order of 0.005 the effect of this factor is small, and it is found that if Q in (48) is replaced by the function $P(f_\alpha/f; a)$ given in figure 5, the results obtained are less than 1% too large.

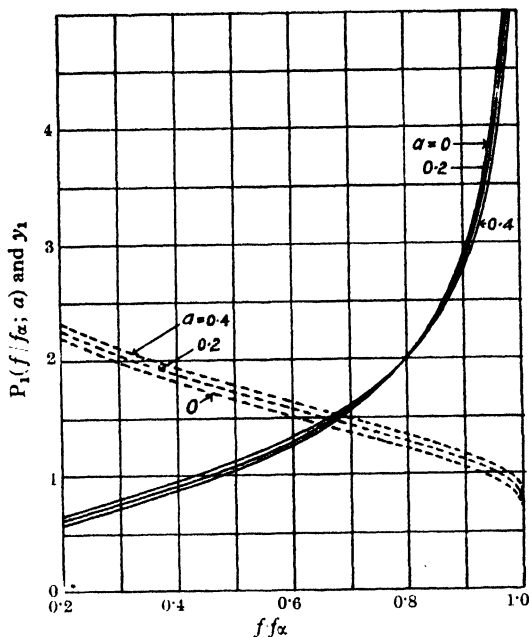


Figure 6.

— Function $P_1(f/f_\alpha; a)$ giving the excess equivalent path for reflection.
 - - - - Function y_1 giving the level of reflection.

For a ray reflected from the layer, the excess of the projection of its path on the earth's surface over twice the projection of the distance measured along OA to the level of reflection is

$$(H/\beta^2) \sec I \sin \alpha Q_1(f/f_\alpha; a; 2H/\beta R), \quad \dots\dots (50)$$

where

$$Q_1\left(\frac{f}{f_\alpha}; a; \frac{2H}{\beta R}\right) = 4 \int_{y_1}^{\infty} \left\{ \frac{1}{(1 - a \ln y - dy \exp(-y^2))^{\frac{1}{2}}} - \frac{1}{(1 - a \ln y)^{\frac{1}{2}}} \right\} \frac{dy}{y[1 - (2H/\beta R) \ln y]^2}, \quad \dots\dots (51)$$

and y_1 is defined in (32) and plotted in figure 6.

As before, Q_1 may be replaced approximately by the function $P_1(f/f_\alpha; a)$ of figure 6; values obtained in this way will be too small, the error being less than 3%.

REFERENCES

- APPLETON, E. V., 1937, *Proc. Roy. Soc. A*, **162**, 451.
 APPLETON, E. V., and BEYNON, W. J. G., 1940, *Proc. Phys. Soc.*, **52**, 518; 1947, *Ibid.*, **59**, 58.
 CHAPMAN, S., 1931, *Proc. Phys. Soc.*, **43**, 26.
 JAEGER, J. C., 1947, *Proc. Phys. Soc.*, **59**, 87.
 MARTYN, D. F., 1935, *Proc. Phys. Soc.*, **47**, 323.

On a New Type of Rotating Anode X-Ray Tube

By A. TAYLOR

Research Department, English Electric Co. Ltd.*

* Now Head of the Physics Department, Mond Nickel Co. Ltd., Birmingham

*Communicated by Sir Lawrence Bragg; MS. received 11 December 1947
 and in amended form 19 February 1948*

ABSTRACT. A brief account is given of the physical limitations which influence the power output of stationary and rotating anode x-ray tubes used in crystallographic work. Brief mention is made of various tubes which incorporate rotating vacuum seals in the anode assembly. An account is then given of the "Peristron"†, a continuously water-cooled rotating anode x-ray tube which completely dispenses with the need of rotating vacuum seals. Incorporated in the tube is a novel type of cathode assembly which permits of rapid adjustment to the focus from outside the vacuum envelope.

§1. REQUIREMENTS AND LIMITATIONS ON X-RAY TUBES FOR CRYSTALLOGRAPHIC PURPOSES

IN registering x-ray diffraction patterns from crystalline and amorphous materials it is not uncommon to require exposure periods lasting many hours. Several factors play their part in determining the exposure time, only a few of which are under the control of the investigator. To a large extent the exposure time is controlled by the nature of the irradiated specimen and on the type of diffraction pattern required.

† From the Greek "peristrophe" (spinning round like a top) and "electron".

Broadly speaking, there are two direct ways in which the exposure times could be reduced by a considerable amount. The first is to increase the speed of the recording x-ray film, and, if this could be done without introducing undesirable large grain-size effects in the emulsion, it would be an achievement of the highest importance. The second method, which is the subject of this communication, is to increase the intensity of radiant energy at the source. This is the most immediately practical method of obtaining a substantial reduction in exposure period.

With the exception of the Seemann-Bohlin type focusing cameras using curved crystal monochromators or Soller slits, the bulk of x-ray diffraction work is carried out with cameras utilizing pinhole apertures or with powder cameras using narrow slits in their collimator systems. These pinholes or slits must be of sufficient size to bathe the specimen completely in radiation, and, at the same time, confine the beam to prevent unwanted scattering effects. In general, the x-ray beam is admitted into the collimator system through a pinhole aperture, or slit, which seldom exceeds 1 mm. across, and which, in special circumstances, may be as small as 0.25 mm. or less. This means that our first major problem in the design of the x-ray tube is to obtain a focal spot of high intrinsic brilliance and with dimensions comparable with the diameter of the pinhole aperture. Our second problem is to position the focal spot as close to the window of the x-ray tube as possible in order to cut down its distance from the irradiated specimen.

Any increase of focal spot dimensions beyond the size of the pinhole aperture merely represents so much waste of x-ray energy. Thus a tube with a nominal electrical energy consumption of 1 kw. is no better for our special purpose than a tube rated at $\frac{1}{2}$ kw. if only one half of the focal spot is utilized, or if, by the construction of the x-ray tube, the distance of the focal spot from the camera is unduly large. In the final analysis, the ideal solution would be to reduce the size of the focal spot so that it can effectively replace the first pinhole aperture of the camera, the lateral definition of the x-ray beam now being produced entirely by the specimen and the focal spot together.

Of the electrical energy dissipated in the focal spot little more than one thousandth part is converted into useful x-radiation, the remainder merely serves to heat the anode. Thus it is generally unsatisfactory to attempt to produce an intense beam of radiation from a small circular focus, for the local heating soon melts a crater in the anode. In the majority of sealed-off and demountable x-ray tubes of the hot-cathode type, the focus is rectangular in shape and with the approximate dimensions 10×1.0 mm. Such a focus, when viewed along its major axis at a grazing angle of 6° to the anode face, presents a foreshortened area of high intrinsic brilliance approximately 1.0 mm. square.

An x-ray tube with water-cooled copper anode and rectangular focal spot as described above, will normally be operated in the region of 40 kv. with an electron current of 20 ma. This total load of 0.8 kw. may be sustained indefinitely without damage to the anode, although further considerations with regard to the life of the filament may necessitate a somewhat lower tube current. This is specially true in the case of sealed-off tubes.

The energy dissipated in the focal spot is thus in the neighbourhood of 0.08 kw/mm². Any effort to increase this loading by concentrating the focal spot into a still smaller area, or by increasing the electrical input, merely raises

the surface of the anode to an unduly high operating temperature. In a tube which has been carefully degassed so that there is no instability from ionization currents, the temperature limit would appear to be determined by the melting point of the anode material. In point of fact, the practical limit is set much lower by the recrystallization of the anode material and the fluctuating thermal stresses to which it is subjected. For the case of copper, de Graaf and Oosterkamp (1938) give an upper limit of 300° c.

It was realized very early in the history of x-ray tubes that a substantially larger output of radiation would be possible if the anode were set in motion thereby continuously exposing a fresh, cool region to the electron bombardment (Breton 1897). This has been achieved by oscillating the target with a rotary movement of a few degrees (Hosemann 1939) and by a slow translatory oscillation of several centimetres (Astbury and MacArthur 1945). It has also been achieved by a gyratory oscillation which gives the effect of continuous rotation (Du Mond and Youtz 1937) and by a true continuous rotation of the anode (Coolidge 1915, 1930, Elihu Thomson 1914, Bouwers 1929). Most ingenious of all is the method in which the entire x-ray tube rotates about the cathode-anode axis while the electron stream, deflected by an external magnet, strikes the anode at a fixed position in space (Breton 1897, Coolidge 1917). Few of these arrangements, are, however, satisfactory for x-ray crystal analysis work.

Müller (1927, 1929, 1931, 1939) has made a detailed mathematical analysis of the heat dissipation in water-cooled stationary and rotating x-ray anodes. Within certain limits, he has shown that for a line focus

$$W_{\max}(\text{rest}) = 4.25(T - T_0)k\lambda(a, b), \quad \dots\dots(1)$$

$$W_{\max}(\text{moving}) = \frac{4.04(T - T_0)ka\sqrt{(vb\rho c/k)}}{1 + (2a/\pi^2 R)(1 + \ln 4R/a)\sqrt{(vb\rho c/k)}}, \quad \dots\dots(2)$$

$$\simeq 4.04(T - T_0)ka\sqrt{(vb\rho c/k)}, \quad \dots\dots(3)$$

where $W_{\max}$ = limiting input in watts, T = maximum safe temperature of surface T_0 = temperature of cooling metal-water interface, a, b = half length (width) of focal spot—assumed elliptical, k = thermal conductivity of target material (watts cm./°c.), ρ = density of target material, c = specific heat of target material (watt.sec./gm./°c.), v = relative velocity of focal spot (cm./sec.), R = radius of path of focal spot on target, $\lambda(a, b) = \lim_{n \rightarrow \infty} A_n = \lim_{n \rightarrow \infty} B_n$, where $A_n = \frac{1}{2}(A_{n-1} + B_{n-1})$ and $A_0 = a, B_0 = b$.

Thus, by giving the anode the velocity v relative to the focal spot, we can increase the performance of the x-ray tube in the ratio

$$\frac{W_{\max}(\text{moving})}{W_{\max}(\text{rest})} = \frac{0.951a\sqrt{(vb\rho c/k)}}{\lambda(a, b)\{1 + (2a/\pi^2 R)(1 + \ln 4R/a)\sqrt{(vb\rho c/k)}\}} \quad \dots\dots(4)$$

In the case of a copper anode, $\rho c/k \simeq 1$. For a focus of dimensions $2a = 0.8$ cm, $2b = 0.05$ cm, $\lambda(a, b) = 0.15$. With an anode of radius $R = 9.0$ cm. which rotates at 1370 r.p.m., we easily find that $W_{\max}(\text{moving})/W_{\max}(\text{rest}) = 11.3$. Equation (4) from which the value 11.3 is derived assumes, among many things, that the x-ray tube is operated at a steady potential. It does not follow then, that if a.c. is applied across the tube the same increase in efficiency is to be expected: in practice, the output is raised by a factor more nearly approaching half the theoretical limit.

§ 2. TYPES OF ROTATING ANODE TUBE

In the case of water-cooled rotating anode tubes we are faced with the dual problem of rotating the anode within the vacuum envelope, and, at the same time, introducing a continuous flow of cooling water into the anode. A number of solutions to this problem have been described in the literature.

The first water-cooled rotating anode tube, developed by Stintzing (1926), was designed primarily for x-ray spectroscopy and was not suitable for crystallographic work. The x-ray tubes developed by Müller and Clay (Müller 1929 b, Clay 1934, Müller and Clay 1939) use a cooling system closely resembling that of Stintzing. The anode is in the form of a flat hollow cylinder driven by a hollow axle which is coupled to an electric motor and rotated at 2000 r.p.m. The flat face of the cylinder, on the side opposite the driving axle, serves as the effective anode surface, the focus marking out a continuous narrow track near its periphery. Cooling water is injected through a co-axial inlet tube housed within the hollow driving shaft and is forced against the inside wall of the anode periphery and then returns by way of the hollow shaft to the outlet pipe. Because of its fundamental importance the arrangement in Clay's 5 kw. x-ray generator is shown diagrammatically in figure 1.

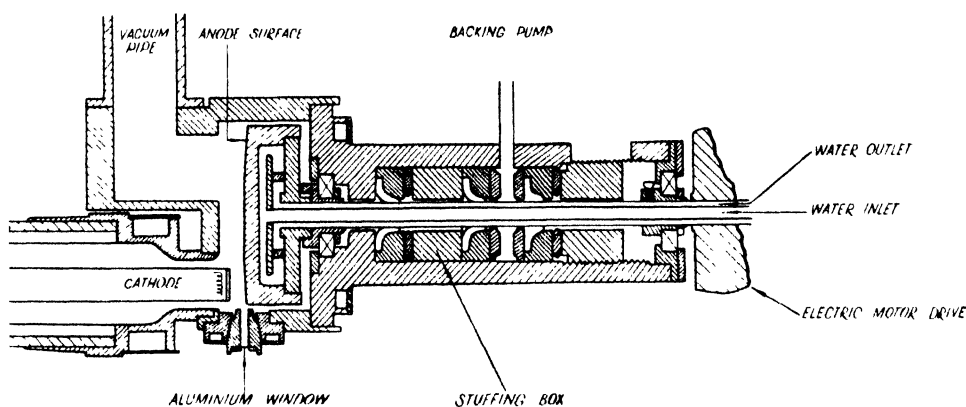


Figure 1. Rotating anode x-ray tube with stuffing box vacuum seal. (Clay.)

Perhaps the most important part of the assembly is the vacuum seal surrounding the rotating shaft. This consists of a stuffing-box packed with hat-leathers pre-treated with vacuum grease and backed with high-speed vacuum pumps. A very similar arrangement is described by Beck (1939).

A stuffing-box type of seal has a number of drawbacks. It leaks when the backing pumps are shut down, and the frictional forces, which produce an appreciable wear on the driving shaft, must be overcome by employing a 2 H.P. electric motor.

These drawbacks led Astbury and Preston (1934) to make use of an almost frictionless permanent liquid seal. This takes the form of a column of mercury of barometric height surrounding the hollow driving shaft of the anode. A detailed description of the apparatus and its performance is given by MacArthur (1944).

A satisfactory and compact seal is provided by a greased conical vacuum joint on the rotating hollow anode shaft. (Fournier, Gondet and Mathieu

1937). The speed of rotation is relatively small (60 r.p.m.), and this limits the energy output.

A description is given by Linnitzki and Gorski (1936) of two rotating anode tubes which eliminate the conventional rotating seals on the driving shafts, replacing them with molecular pumps of the combined Gaede-Hollweg or Gaede-Siegbahn type. No performance data for these tubes are given.

From the brief descriptions of the rotating anode x-ray tubes given above, one feature common to them all becomes evident. That is, the cooling water

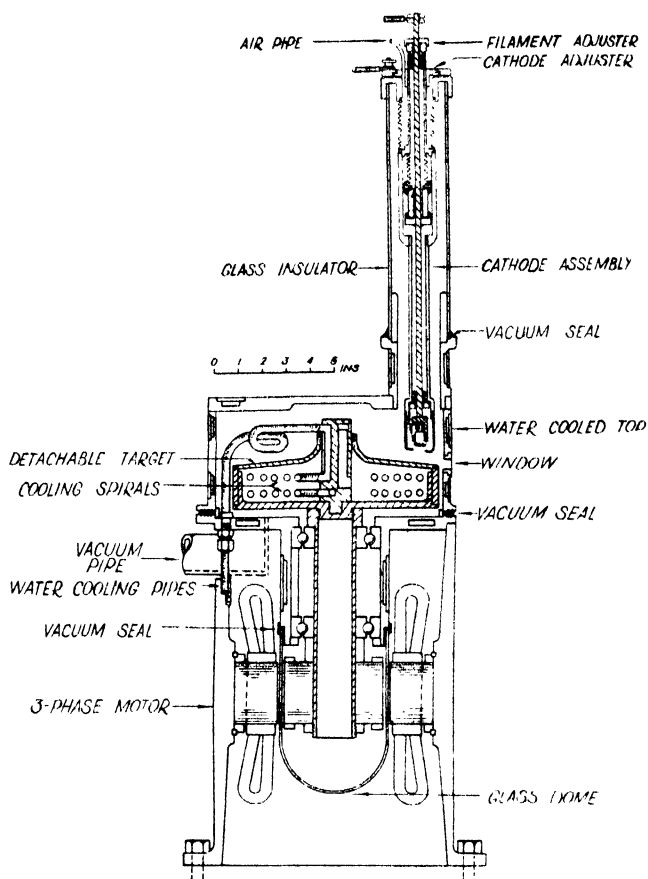


Figure 2. Section through x-ray tube.

is led to the inner surface of the anode through a hollow driving shaft, which, in its turn, must pass the wall of the x-ray tube by way of some form of rotating vacuum joint. The Peristron (Taylor 1944) described below, is a rotating anode x-ray tube using continuous water cooling, and eliminating entirely the need for rotating vacuum seals.

§ 3. DESCRIPTION OF THE PERISTRON

Figure 2 is a vertical section through the rotating anode x-ray tube. Motive power to the rotating anode is given by a 3-phase 400-volt induction motor having a specially large air-gap $5/32''$ across and therefore consuming only a

small fraction of 1 H.P. The speed of rotation is 1370 r.p.m. The brass base of the vacuum envelope of the x-ray tube is accurately spigotted into the cylindrical top of the induction motor which serves as a very rigid support. Brazed into the base is a right-angled vacuum pipe $1\frac{1}{2}$ " in diameter. The centre of the base houses the ball-races for the driving shaft in a central tubular member, the lower end of which is terminated by a waxed-on $\frac{1}{8}$ " thick glass dome. This occupies the air-gap of the motor and effectively seals the rotor, at the lower end of the driving shaft, within the vacuum envelope.

Screwed into the upper end of the shaft is the detachable hollow anode. The whole sub-assembly, comprising anode, base-plate, rotor and glass dome, removed from the motor housing, is shown in figure 3.

The water-cooled brass top of the x-ray tube is vacuum sealed to the base-plate by means of a flat neoprene gasket. The flanges gripping the gasket in position are turned with a series of fine concentric ridges. These bite into the neoprene so that the atmospheric pressure bearing down on the top is quite sufficient to maintain a perfect vacuum seal which is both easily demountable and trouble-free. The cathode assembly is mounted on the top by means of a demountable plasticine vacuum joint. The window, which is of lithium metal 1 mm. thick, is pressed into position by a screw cap and is easily replaced.

(i) *The anode assembly*

The anode with its associated cooling system comprise the special feature of the x-ray tube. It is essentially a hollow cylinder, the base of which is of brass $8\frac{1}{2}$ " in diameter and $\frac{1}{4}$ " thick, with a screwed boss on the lower surface for attachment to the driving shaft of the electric motor. The upper anode surface is made of pure copper sheet, $\frac{1}{8}$ " thick and with a slope of 6° , the focal spot delineating a track approximately 10 mm. wide at a distance of $\frac{1}{4}$ " from the periphery. In the original construction, the bush at the centre of the anode surface and the cylindrical walls were of brass and were brazed to the copper anode, but a much simpler construction was evolved in which the bush, anode surface and cylindrical walls were spun from a single copper sheet.

A flat double-spiral of thin-walled copper tubing, $5/16$ " external diameter and $\frac{1}{4}$ " bore is held stationary within the hollow anode by the spindle which passes through the centre bush in the upper surface of the rotating anode. Water is introduced through pipes passing through stationary demountable seals in the base of the vacuum envelope and passes, in turn, through the spindle into the spirals within the anode and out again through the spindle.

The water-cooling leads to the spirals are bridged together by a brass plate which bears on a rubber gasket, thus forming a demountable vacuum seal covering the holes where the nipples on the pipes pass through the base. The bridge also serves to prevent any twisting of the pipes as the locknuts on the vacuum seals and pipe nipples are tightened.

The anode is partly filled with vacuum oil through the large inlet hole drilled in the spindle which is then closed by means of a screwed filler-cap. About $\frac{1}{2}$ litre of oil is sufficient. The surface of the oil and its vapour are in active communication with the high vacuum within the x-ray tube through the gap between the spindle and the bushing. Thus it is necessary to use an oil with an extremely low vapour pressure such as Apiezon "B", Octoil "S" or a silicone. In the present instance Apiezon "B" with its vapour pressure of 10^{-6} mm.

has been successfully used. Any possible creep of oil through the gap between the spindle and bush is prevented by a screw thread cut into the bush which drives the oil back into the anode. A clearance of about $10/1000''$ is required to prevent the bush seizing on the spindle, since the vacuum oil has very poor lubricating properties.

In the course of rotation, the centrifugal forces fling the oil against the circumferential inner surface of the anode. By this means, the heat generated in the metal surface of the anode by the electron bombardment is instantly transferred to the oil, which in its turn is transferred to the cooling water flowing through the stationary spirals. Moreover, the churning motion imparted to the oil as it is forced past the spirals provides an extremely efficient method of transferring the heat, for the cooling is not dependent on natural convection currents as in a static body of liquid. With a flow of water of only 10 litres per minute, and an initial temperature of 14°C. , the outflow temperature was 20°C. when the power dissipated in the anode was 5 kw.

The design of the x-ray tube is such that the anode is easily accessible for cleaning and for replacement whenever different characteristic radiation is required. Upon lifting away the lid from the rubber gasket which forms the vacuum seal with the base-plate, the whole anode assembly is exposed (figure 4). To remove the anode, the water-cooling pipe nipples and locknuts on the base-plate are unscrewed and then the anode is unscrewed from the driving shaft, which is held stationary by means of a flat spanner. The driving shaft, bearings and rotor are thus left behind as an integral part of the base-plate and need never be disturbed. The time required to strip down the tube and replace the anode is little more than 5 minutes.

After re-assembly it is important to rotate the anode before commencing pumping operations. This prevents any tendency for oil to creep past the bush should it foam during the escape of adsorbed air. For this reason a thermal delay switch was incorporated in the electrical controls which gave a few seconds for the anode to rotate up to speed before allowing the rotary pump to start. Once vacuum conditions were established the anode could be rotated or stopped at will independently of the pumps.

The x-ray tube is mounted on a steel stand which houses the vacuum pumps and also serves to carry an electrical control panel. The pumping unit comprises Metropolitan Vickers O2 and O3 oil condensation pumps mounted in cascade and backed by a DR1 type rotary pump. The cooling water from the pumps is arranged to pass into the cooling spirals of the anode. This is a more efficient arrangement than passing the water in the reverse direction since the temperature rise on passing through the pumps is negligibly small. It has the advantage over duplicate sets of cooling leads that only one water relay safety switch is required in the electrical circuit.

Directly beneath the platform supporting the x-ray tube are two small air compressors. One circulates a stream of cooling air round the stator windings and glass dome of the x-ray tube, the other serves to cool the cathode assembly.

(ii) *The cathode assembly*

In the majority of demountable x-ray tubes, the provision made for adjusting the position of filament and focusing tube with respect to the anode is very crude, and any alterations must be made by opening up the vacuum system. This is a time-consuming process with the result that an operator is too often content

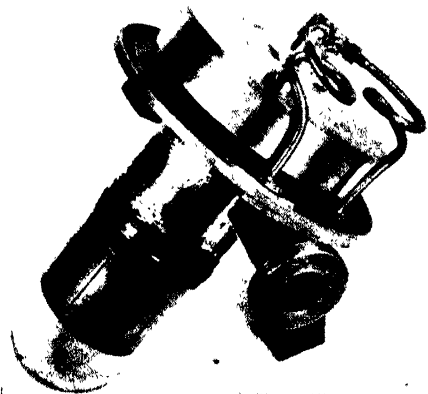


Figure 3. Anode sub-assembly removed from motor-housing.

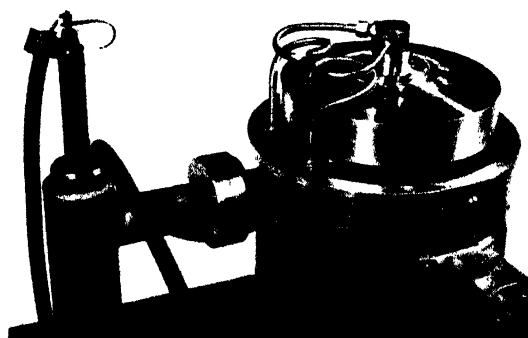


Figure 4. Rotating anode x-ray tube with top removed, exposing anode assembly.

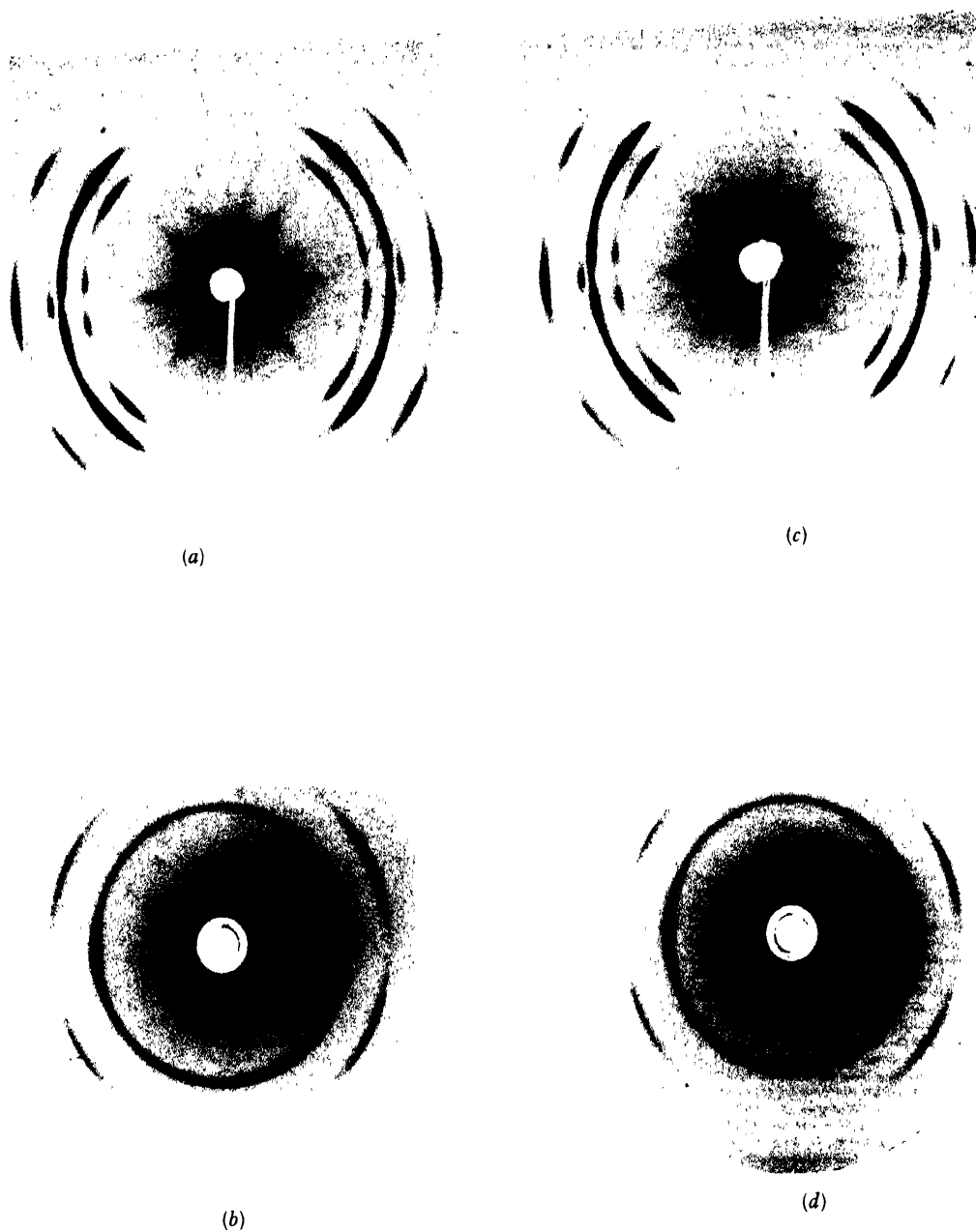


Figure 6. Transmission (a) and back-reflection (b) photographs of 0.0015" copper foil, taken with stationary anode x-ray tube. $\text{CuK}\alpha + \beta$ radiation. 40 kvp. 20 ma. Exposure time 20 minutes. (c) & (d) Rotating anode tube. 40 kvp. 100 ma. Exposure time 3 minutes. (Reproduced $\frac{1}{2}$ size.)

to let the tube run with an inefficient broad focus. Moreover, it is not generally realized that if the tube is focused for running with a chromium anode at 35 kv. a change to 90 kv. for operating with a silver or tungsten anode will have a marked defocusing effect. It was in order to overcome these difficulties that a new adjustable type of cathode assembly was developed.

In the present arrangement shown in figure 5, the position of the cathode assembly relative to the anode is altered by means of a flexible metal bellows moved by a nut, the cathode adjuster, situated outside the vacuum system on the metal cap sealed to the glass insulator. A second internal bellows is connected to the filament holder, which may be moved up or down within the focusing tube by means of the filament adjuster nut, which is soldered to a porcelain insulator above the cathode adjuster. It is therefore possible to make adjustments and obtain the best focusing conditions in a few seconds by observing the changes in a pinhole image of the focal spot upon a fluorescent screen—an operation which normally takes several hours with the more orthodox type of cathode arrangement.

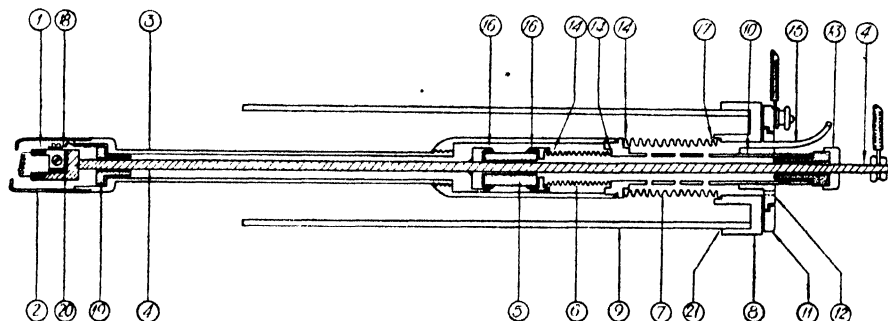


Figure 5. Cathode assembly.

LIST OF PARTS

- | | | |
|------------------------|------------------------------------|----------------------------------|
| 1. Filament assembly | 9. Glass insulator | 16. Metal-porcelain vacuum joint |
| 2. Focusing tube | 10. Screwed tube | 17. Bellows-cap vacuum joint |
| 3. Steel tube | 11. Clamping ring | 18. Flexible filament lead |
| 4. Filament lead | 12. Cathode assembly adjusting nut | 19. Porcelain guide |
| 5. Porcelain insulator | 13. Filament retractor nut | 20. Mica insulation |
| 6. Flexible bellows | 14. Soldered vacuum joint | 21. Wax vacuum seal |
| 7. Flexible bellows | 15. Air pipe | |
| 8. Brass cap | | |

Experiments were made with a number of filament sizes and types of focusing tubes. Ultimately a simple combination was found which gave a sharp focus of 10 mm. $\times$ 0.5 mm. at 45 kv., and 100 ma. This consisted of a tungsten spiral of 16 turns of 0.33 mm. diameter wire wound on a mandril 0.8 mm. diameter into a helix 10 mm. long, housed within a steel focusing tube with a simple slot 2½ mm. across and 12 mm. long. The slot opens out from the surface nearest the anode towards the filament at an angle of 45°. The focusing tube stands 15 mm. from the target with the filament surface approximately 1 mm. behind the front surface of the slot. This latter distance is most critical, for a movement of $\frac{1}{10}$ mm. in either direction will produce considerable changes in the sharpness of focus. With the target bevelled at 6° the effective focal spot is a rectangle 1 mm. high by 0.5 mm. across. A filament current of 8.5 amp. is sufficient for a tube current of 100 ma. Excellent results have also been obtained with tantalum filaments.

The filament is held in position in the simplest possible manner. The straight ends of the filament spiral are pushed into 1 mm. diameter holes drilled

in the filament holder and held rigidly in position by driving in thin soft iron wedges. These have the great advantage of being easily removed when the filament is due for replacement. Since an appreciable amount of heat is conducted from the filament along the steel cathode assembly to the soldered joints on the bellows and to the Apiezon "W" wax seal between the glass insulator and the end cap, some form of cooling is desirable. A stream of cooling air, generated by a small compressor mounted below the x-ray tube, is circulated through the bellows system via an inlet pipe soldered into the cathode adjuster nut. Such cooling only becomes necessary if the tube is to be run continuously for periods of over an hour.

§ 4. PERFORMANCE OF THE X-RAY TUBE

The x-ray tube may be operated continuously at 100–120 ma. and 45 kv. without giving any trouble. It has also been run for varying periods at 30 kv. at 175 to 200 ma. It must be stated that these figures are nominal in that raw A.C. is applied from the H.T. transformer directly to the tube and the resultant D.C. component in the transformer may cause serious distortion of the wave form. In this work the focal spot was 0.5×10.0 mm.

As a basis of comparison, photographs were taken with the same specimen in the same camera with the Peristron and with a new sealed-off Victor x-ray tube run at its maximum output. Figure 6 illustrates transmission and back reflection photographs of $1\frac{1}{2}/1000''$ copper foil taken with the two tubes using unfiltered $\text{Cu } \alpha + \beta$ radiation. The reduction in exposure time is from 20 minutes to 3 minutes. On account of its extremely sharp focal spot, the rotating anode tube gives slightly sharper photographs.

ACKNOWLEDGMENTS

The author wishes to thank the English Electric Co. Ltd., in whose research laboratories the Peristron was developed, for their kind permission to publish this paper, and Mr. J. K. Brown, Director of Research, for his very helpful interest. He also wishes to acknowledge the great assistance given by Mr. F. Leighton in the construction and operation of the equipment.

REFERENCES

- ASTBURY, W. T., and MACARTHUR, I., 1945, *Nature, Lond.*, **155**, 108.
 ASTBURY, W. T., and PRESTON, R. D., 1934, *Nature, Lond.*, **133**, 460.
 BECK, J., 1939, *Phys. Z.*, **40**, 474.
 BOUWERS, A., 1929, *Verh. dtsch. Röntgen-Ges.*, **20**, 102.
 BRETON, J. L., 1897, *Les Rayons Cathodiques et Rayons X. Rev. Sci. et Ind.*, p. 86.
 CLAY, R. E., 1934, *Proc. Phys. Soc.*, **46**, 703.
 COOLIDGE, W. D., 1915, *Amer. J. Roentgenol.*, 884; 1917, U.S. Patent No. 1215116; 1930, *Gen. Elect. Rev.*, **33**, 608.
 DU MOND, J. W., and YOUTZ, J. P., 1937, *Rev. Sci. Instrum.*, **8**, 291.
 FOURNIER, CONDET and MATHIEU, 1937, *J. Phys. Radium*, **8**, 160.
 DE GRAAF, J. E., and OOSTERKAMP, W. J., 1938, *J. Sci. Instrum.*, **15**, 293.
 HOSEMANN, R., 1939, *Z. tech. Phys.*, **20**, 203.
 LINNITZKI, V., and GORSKI, V., 1936, *Tech. Phys. U.S.S.R.*, **3**, 220.
 MACARTHUR, I., 1944, *Electronic Engng.*, **17**.
 MÜLLER, A., 1927, *Proc. Roy. Soc. A*, **117**, 30; 1929 a, *Ibid.*, **125**, 507; 1929 b, *Nature, Lond.*, **124**, 128; 1931, *Proc. Roy. Soc. A*, **132**, 646.
 MÜLLER, A., and CLAY, R. E., 1939, *J. Instn. Elect. Engrs.*, **84**, 261.
 STINTZING, H., 1926, *Phys. Z.*, **27**, 844.
 TAYLOR, A., 1944, British Patent Applications 25718/44 and 37511/46.
 THOMSON, ELIHU, 1914, U.S. Patent, described in Coolidge (1930).

LETTERS TO THE EDITOR

The Dipole Moment of the Supposed Fundamental Magnetic Field due to Rotation

In his recent important discussion of the magnetic fields of the earth, sun and stars, Prof. Blackett (1947) suggested that these fields represent a new and fundamental property of rotating matter. His formulation of his hypothesis (reviving and revising one already proposed by H. A. Wilson) is at present tentative, and applicable only to bodies symmetrical about their axis of rotation; it is that ρv , the mass flux due to the rotation, is associated with a fundamental magnetic field as if it were an electric current intensity $K\rho v$; here ρ denotes the mass density, v the rotating velocity ωp (ω the angular velocity and p the distance from the axis), and K , when the current intensity is expressed in electromagnetic units, is given as $\beta G^{1/2}/2c$ (G the constant of gravitation, c the speed of light, and β a factor of order unity).

As Blackett indicated, the external fundamental field of a *spherical* body in *uniform* rotation, when the density distribution is *spherically symmetrical*, is of dipole character, and the following relation holds between the dipole moment P and the angular momentum U : $P/U = \frac{1}{2}K$. Thus P/U is independent of the nature of the function $\rho(r)$ expressing the spherically symmetrical density distribution.

In this note I show, very simply, that $P/U = \frac{1}{2}K$ for a rotating body when there is symmetry about the axis of rotation, whatever the form of the body, or the non-uniformity either of ρ or ω .

Let the position of any point of a half meridian section (bounded by the axis) be specified by p , the distance from the axis, and z , the coordinate measured parallel to the axis. Let $dp dz$ be the area of an element of the section, with sides parallel and perpendicular to the axis, at the point p, z . The equivalent electric current associated with this element is $K\rho\omega p dp dz$; it flows round a circle of radius p . By Ampère's hypothesis its field is the same as that of the magnetic induction B for a uniform magnetic shell of strength $j = K\rho\omega p dp dz$ bounded by the circle; it is convenient to take this shell as plane; its magnetic moment is j times its area (πp^2) and is parallel to the axis. As moment vectors are additive, and as these moments are all in the same direction along the axis (except in so far as ω may vary in sign), the total dipole moment P is given by

$$P = \sum \pi p^2 j = \iint \pi K \rho \omega p^3 dp dz.$$

In this expression ρ and ω may be any functions of p and z .

The mass of the ring associated with the element $dp dz$ is $2\pi p \rho dp dz$, and its speed is ωp ; its angular momentum is therefore $2\pi \rho \omega p^3 dp dz$, and thus by integration $U = 2\pi \int \rho \omega p^3 dp dz$. Hence $P/U = \frac{1}{2}K$, as stated.

If C is the centre of the current ring associated with the element $dp dz$ (so that C is the axial point with coordinate z), the potential of the field of the ring, at a distant point, can be expressed as a series of zonal harmonic terms, of which the first is determined by the dipole moment of the ring. If the potential is expressed relative to any other point O on the axis, this first term is unaltered, but the coefficients of all the higher harmonics are changed. When the contributions of all the rings are combined to give the whole potential, the first harmonic corresponds to the dipole moment $\frac{1}{2}KU$, and in addition there are in general higher harmonics which do not vanish whatever the choice of O . These do vanish, however, for a spherical body rotating about a diameter, when ρ and ω are functions only of the radial distance.

In a paper on "Solar magnetism and the hypothesis of fundamental magnetization by rotation", to be communicated to the Royal Astronomical Society, some examples of these higher harmonics are given for a non-uniformly rotating sun.

Queen's College, Oxford.
26th April 1948.

S. CHAPMAN.

The Freezing-in of Nuclear Equilibrium

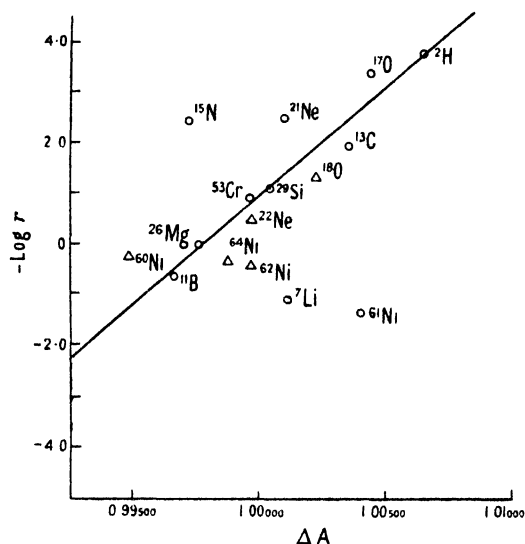
In a previous paper (Ubbelohde 1947) a method was described for testing whether the abundance ratios of stable isotopes might be due to the freezing-in of thermodynamic equilibrium. The test was constructed so as to avoid uncertainties about the density of the system, and about the mass of the neutron, at the time of freezing-in. From the freezing-in curve a freezing-in temperature was calculated of approximately 10^{10} °K.

Further data have since become available (Mattauch 1946, Stephens 1947). Where the masses of the stable isotopes are known to at least four places of decimals, it is possible to test whether the additional abundance ratios are in agreement with the freezing-in curve. Uncertainties in ΔA greater than 0.0002 prevent any decision for all the elements with $Z > 28$. No conclusions can yet be reached about their non-thermodynamic origin, on account of this gap in the data.

New values of ΔA of sufficient accuracy for the test include the following :

$m+1A_n$	Abundance ratio ($m+1$)/ m	ΔA (mass units)	$m+2A_n$	Abundance ratio ($m+2$)/ m	ΔA (mass units)
^3He	10 ⁷	0.986872	^{18}O	0.0020	2.00485
^{11}B	4.00	0.996732	^{60}Ni	0.396	1.99010
^{13}C	0.011	1.003681	^{62}Ni	0.142	1.99978
^{15}N	0.00384	0.997340	^{64}Ni	0.221	1.99785
^{17}O	0.0004	1.00450			
^{61}Ni	0.0449	1.00419			

When these values are plotted on the curve of $-\log r$ against ΔA , previously established (Ubbelohde 1947) (see figure), it will be seen that for the isotopes ^{11}B , ^{13}C , ^{17}O , ^{18}O , ^{60}Ni , ^{62}Ni , ^{64}Ni there is good agreement within the limits of uncertainty previously found. (To avoid confusion, not all the earlier data are included in the figure.) For the isotopes ^3He , ^{15}N and ^{61}Ni no thermodynamic relation can be found, unless the reactions maintaining equilibrium are very notably slower and faster respectively than the average.



Activity of the neutron at freezing-in

The value of the intercept made by the freezing-in curve when $\log r = 0$ is $\Delta A = 0.99780$. This intercept may be used in conjunction with the freezing-in temperature and an estimated value for the mass of the neutron, to calculate the neutron activity at the time of freezing-in.

For convenience, the equilibrium equations of the previous paper are rewritten :

$$-\log n + \log r = -4.7 \times 10^{12} (\Delta A - m_n) / T + 3/4.606 - \frac{3}{2} \log T - \log [(2\pi m_n k)^{1/2} k / h^3], \dots (1)$$

where the effects of the difference in mass between isotope pairs, and any differences in spin multiplicities have been neglected in the last term.

Writing $\log r=0$, $T=10^{10}$, $\Delta A=0.99780$, $m_n=1.008941$ (cf. Stephens 1947), we find $-\log n=-13.87$. There seems to be no likelihood of a value of n as large as this. The numerical result is, however, very sensitive to the exact numerical value taken for the freezing-in temperature. It is rather less sensitive to the value assumed for the mass of the neutron. For example, if a freezing-in temperature of 10^9 were used, $-\log n=+34.9$.

This value of n is very much too small to be probable. Clearly a temperature intermediate between 10^{10} °K. and 10^9 °K. would give quite plausible values of n . But the test for the thermodynamic equilibrium is best constructed so as to avoid the use of n altogether, as indicated in the previous paper.

It is important to note that the very rapid change in neutron concentration over the temperature range between 10^{10} °K. and 10^9 °K. is imposed by the physical magnitudes in equation (1). This rapid change provides a simple explanation for the suggestion made in the previous paper (Ubbelohde 1947) that freezing-in occurred on account not of a deficiency in activation energy, but of a deficiency in neutron concentration at around the freezing-in temperature.

With regard to the effect of the mass of the neutron, using the values given above $\Delta A - m_n = -0.01114$. The estimated uncertainty in m_n is ± 0.00002 , which would not affect the calculations appreciably. The maximum uncertainty in the ΔA intercept is about ± 0.00150 , and this would not be enough to make a serious difference in n . But at the temperature of 10^{10} °K., the R.M.S. velocity of particles of unit mass ($^{16}\text{O}=16$) is approximately 1.5×10^9 cm/sec., i.e. 0.05β . Neutrons hotter than the average will begin to exhibit relativity effects, which may to some extent modify equation (1).

Thanks are due to Mr. G. E. Roberts and Mr. J. K. Brown for drawing my attention to the additional data quoted above.

Queen's University, Belfast.

A. R. UBBELOHDE.

25th March 1948.

MATTAUCH, J., 1946, *Nuclear Physics Tables* (Interscience Publ. Inc.).

STEPHENS, W. E., 1947, *Rev. Mod. Phys.*, **19**, 19.

UBBELOHDE, A. R., 1947, *Proc. Phys. Soc.*, **59**, 139.

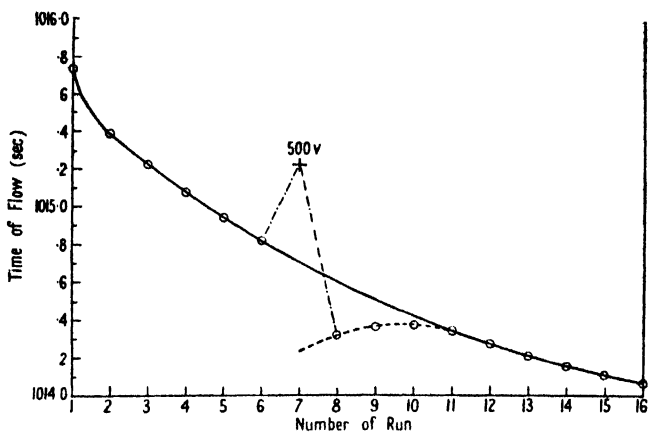
The Detection of Heating in Liquid Dielectrics at Low Frequencies

The heating in dielectrics which occurs at radio frequencies is well known and is so marked that its detection presents no experimental difficulty. It has indeed found many applications in industry since the uniform production of heat throughout the substance is difficult, if not impossible, to produce by other methods. The heating is very dependent on the frequency of the alternating electric field, and at low frequencies, say 10^4 c/sec., becomes so small that its detection and accurate measurement are extremely difficult.

During the course of some experiments by Prof. E. N. da C. Andrade and myself on the effect of alternating electric fields on the viscosity of certain liquids, the dielectric heating of certain polar liquids not only showed itself clearly but proved to be capable of rough quantitative estimation at frequencies as low as 10^4 c/sec. In these experiments a viscometer is used in which the liquid flows through a narrow rectangular channel between massive steel plates (Andrade and Dodd 1946). The viscometer is contained in an air bath immersed in a large water bath whose temperature is controlled so that its variation is less than $1/1000$ degree over a period of many hours. The time of flow of a certain volume of the liquid through the channel is measured by the successive interruptions of two beams of light by the liquid falling in one limb of the viscometer. The amplified electric impulses from the photoelectric cells upon which the light beams are directed are applied to high-speed relays used to start and stop a synchronous electric clock reading to 0.001 second. Times of flow taken at four-hourly intervals do not differ by more than 0.01 second for a total time of flow of over 1000 seconds, showing that evaporation is negligible.

Alternating electric fields of amplitude 50 kv/cm. and frequencies up to 20,000 c/sec. have been applied to the liquid whilst it was flowing through the gap, the times of flow of the liquid being measured not only while the field is applied but also both before and after the application of the field.

With monochlorobenzene, times of flow taken immediately after one another are plotted in the figure, the field of amplitude 35 kv/cm., and frequency 10^4 c/sec., being applied after six measurements without field. The steady decrease of the first six values has been shown to be due to the slight heating of the liquid caused by the dissipation of the work done in forcing it through the capillary. A much bigger decrease occurs immediately after the application of the field, indicating that a new form of heating has occurred. This heat is



slowly absorbed by the thermostat, and after four or five runs the times fall accurately on the curve through the points obtained before the application of the field. This can only be dielectric heating as the liquid is an almost perfect insulator, so that the ordinary Joule heating is absent. Confirmation of this arises from the fact that no such decrease in time of flow was observed immediately after the application of a *steady* field of 40 kv/cm. The increase in the time of flow which occurs whilst the field is being applied is, of course, due to the increase in the viscosity of the liquid and will be fully discussed elsewhere.

The dielectric heating which occurs whilst the field is being applied can be obtained by extrapolation and is seen to cause 0.46 sec. decrease in the time of flow of 1015 sec. With the liquid in question this represents a mean temperature rise of about 0.04° .

The purpose of this note is to draw attention to this method of detecting very small dielectric heating in liquids. The thermal capacity and surface area of the electrodes, although appropriate for the main purpose for which they were designed, are too large for the exact estimation of the dielectric heating. However, a rough estimation can be made of the power factor at these low frequencies on the following considerations.

The heat generated in the liquid will be communicated to the steel plates forming the viscometer. The velocity of propagation of a simple harmonic temperature wave of period T through the metal of diffusivity h is $\sqrt{(4\pi h/T)}$. Taking h for steel as $0.15 \text{ cm}^2 \text{ sec}^{-1}$ and T as 2000 sec., the value of this velocity is about $0.03 \text{ cm. sec}^{-1}$ and will, of course, be greater for the higher frequency terms in the Fourier series of simple harmonic terms which represents the actual applied temperature wave. This means that the temperature wave will have travelled the thickness of the steel plates in a time of the order of 100 sec. The damping factor $\sqrt{(\pi/hT)}$ is small in this case, being approximately 0.1, so that inside the steel at 1 cm. from the liquid/steel boundary the temperature excess is 9/10 of the value at the interface. Hence, during the time of flow with the electric field applied, the whole of the metal, of thermal capacity $260 \text{ cal. deg}^{-1} \text{ c.}$, may be considered to a first approximation as being at the temperature of the liquid. The heat which must be given to the liquid and the surrounding metal in order to produce the rise of temperature noted (0.04° c.) is thus 10.8 calories.

During the flow the whole surface of the apparatus is losing heat to the air bath, whose temperature is 0.04° c. lower. Taking the surface area as 320 cm^2 and its emissivity as $2.5 \times 10^{-4} \text{ cal. cm}^{-2} \text{ sec}^{-1} \text{ deg}^{-1} \text{ c.}$, the heat given to the enclosure is about 3.2 calories, so that the total quantity of heat generated in the liquid is 14.0 calories in 1015 sec. Since the volume of liquid present at any instant in the capillary is 0.143 cm^3 , the mean rate of production of heat in unit volume of liquid is $9.6 \times 10^{-2} \text{ cal. cm}^{-3} \text{ sec}^{-1}$.

Theoretically the condenser with dielectric may be considered as made up of a pure capacity in series with a small pure resistance. The rate of dissipation of energy W , per unit volume of dielectric, when an alternating field of amplitude X and frequency ν is applied to the system is then $\frac{1}{2} X^2 K \nu \cos \phi$, where K is the dielectric constant and $\cos \phi$ is the power factor. In the case considered, $X = 35 \text{ kv. cm}^{-1} = 1.17 \times 10^2 \text{ E.S.U. cm}^{-1}$, $\nu = 10^4 \text{ sec}^{-1}$ and $K = 6$, so that $W = 4.9 \cos \phi \text{ cal. cm}^{-3} \text{ sec}^{-1}$,

Equating this expression for W to that obtained by experiment gives the value 2×10^{-3} for the power factor for chlorbenzene under the conditions holding in the experiment.

Similar effects have been observed with higher fields and frequencies for monochlorbenzene, chloroform and amyl acetate, all of which are polar.

No change in time of flow greater than 1 part in 100,000 was observed with benzene. This liquid is non-polar, and any dielectric heating can only be obtained through the distortion polarization of the molecule, and would, therefore, be expected to be very small, since the power factor is, in this case, of the order of 2×10^{-4} .

Physics Dept., University College, London.
20th April 1948.

C. DODD.

ANDRADE, E. N. DA C., and DODD, C., 1946, *Proc. Roy. Soc. A*, **187**, 299.

The B-N Band-system of Silver Iodide

In the course of work on the absorption spectrum of silver-iodide vapour in the far ultra-violet region (Barrow and Mulcahy 1948) it was noticed that within a rather narrow range of pressure the centre of the B-N system at 3200 Å. was well developed. The long wavelength end of this system has been the subject of careful investigation by Brice (1931), but his measurements end with the 0,4 band at 3295 Å., and in the region 3170–3295 Å. only the rather rough data of Franck and Kuhn (1927) are available. We have therefore thought it useful to record our measurements, especially since they lead to somewhat different values of ω'_0 and D' from those given by Brice.

The bands were photographed in a first order of a 2.4-m. grating instrument on Ilford Process plates, the linear dispersion on the plate being about 7.4 Å/mm. A tungsten filament lamp with "Cornex" envelope, kindly supplied by the Research Laboratories of the G.E.C., was used as source of continuum. The optimum temperature of silver iodide was about 900° C. Of the sixteen bands measured, only for the 2,4 and 0,3 bands was the silver isotope separation clearly resolved: for the other bands it was assumed that settings on the low- λ side of the origin were on ^{107}AgI heads and on the high- λ side on ^{109}AgI heads.

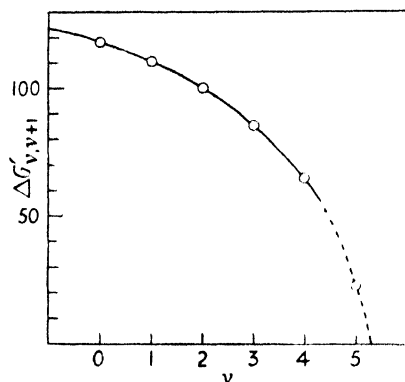
Table 1. Band-head data for ^{107}AgI

6	31657 (0)								
	<i>3157.9</i>								
	(23)								
5	31633.6 (1)								
	<i>3160.28</i>								
	64.9								
4	31568.7 (2)								
	<i>3166.78</i>								
	86.4								
3	31482.3 (3)						30868.4 (1)		
	<i>3175.47</i>						<i>3238.63</i>		
	98.2						99.6		
2	31384.1 (7)	206.9	31177.2 (3)				30768.8 (2)	200.8	30568.0 (2)
	<i>3185.41</i>		<i>3206.55</i>				<i>3249.11</i>		<i>3270.45</i>
	111.3		110.1						
1	31272.8 (6)	205.7	31067.1 (7)	205.8	30861.3 (3)				
	<i>3196.75</i>		<i>3217.91</i>		<i>3239.37</i>				
	118.4		119.0		116.7				
0	31154.4 (5)	206.3	30948.1 (9)	203.5	30744.6 (10)	204.2	30540.4 (8)		
	<i>3208.90</i>		<i>3230.29</i>		<i>3251.67</i>		<i>3273.41</i>		
	0		1		2		3		4
	$\nu'' \rightarrow$								

Italic figures denote wavelengths in air; roman, wave numbers *in vacuo*; figures in parentheses are visual relative intensities.

A correction was applied to the latter, so that all the figures in table 1, which summarizes the band-head data, relate to ^{107}AgI .

The results, particularly for the $v''=0$ progression, provide new information about the vibrational levels of the excited state. This is given in table 2 in the form of energies of levels v' in excess of $v''=0$ (for ^{107}AgI), calculated both from the bands in table 1 and from Brice's figures for ^{107}AgI and ^{109}AgI . In the calculations, Brice's expression for G_v'' has been used. The agreement is very satisfactory, and shows that the expression for G_v'' deduced for bands with v'' ranging from 4 to 17 holds well over the range 0 to 17. From the mean results of table 2, values of $\Delta G_{v, v+1}'$ were determined and plotted as in the figure.



Course of the vibrational intervals in the B state of ^{107}AgI .

We obtain $\omega'_e=123.5$, $D'_e=570$, $D'_0=510$, which differ significantly from the figures given by Brice, namely, $\omega'_e=131.3$, $D'_e=769\text{ cm}^{-1}$.

Table 2. Energies of levels v' in excess of $v''=0$ (cm^{-1})

v'	Present work	Brice	Mean	$\Delta G'$
6	31657 (1)*	—	(31657)	
5	31633.6 (1)	—	31633.6	(23)
4	568.7 (1)	31568.6 (4)*	568.6	65.0
3	482.0 (2)	482.9 (14)	482.8	85.8
2	383.3 (5)	382.3 (12)	382.6	100.2
1	272.1 (3)	272.0 (8)	272.0	110.6
0	153.9 (5)	153.9 (3)	153.9	118.1

* Number of band-heads from which the values were derived

The unusual character of the B state (Mulliken 1937) is emphasized by these results. Thus, from the distribution of intensity among the bands, it appears that r'_e is somewhat, but not very much, greater than r''_e ; yet the force-constant in state B is only 36% of its value in the ground state, while the vibrational levels converge so rapidly that D'_0 is only 3% of the dissociation energy of the ground state.

Physical Chemistry Laboratory,
Oxford.

R. F. BARROW.
M. F. R. MULCAHY.

22nd April 1948.

BARROW, R. F., and MULCAHY, M. F. R., 1948, *Nature, Lond.*, in the press.

BRICE, B. A., 1931, *Phys. Rev.*, **38**, 658.

FRANCK, J., and KUHN, H., 1927, *Z. Phys.*, **43**, 164.

MULLIKEN, R. S., 1937, *Phys. Rev.*, **51**, 310.

DISCUSSION

on paper by R. WEIL entitled "The Variation with Temperature of Metallic Reflectivity" (*Proc. Phys. Soc.*, 1948, **60**, 8).

Dr. D. J. PRICE. The equation preceding (15) has the obvious factor of $2x - a$ leading to a root having a value $\frac{1}{2}a$. This root leads to λ_x being of the magnitude 0.1μ for most metals. If, as stated, the cubic had only one real root, the above would be the only solution. This, however, is also false. It is probable that the complicated substitutions and expansions have given stationary points for that portion of the reflectivity curve giving negative values for the reflectivity. (These are, of course, without physical significance.)

From purely physical grounds, the conclusion of (16) may be seen to be unsound, since λ_r is the limit at which the Hagen-Rubens relationship becomes valid to a high approximation, and, therefore, the anomalous behaviour associated with the appearance of an X -point cannot occur here.

Taking all this into account, it would appear then that the only region in which an X -point may be found is in wavelengths small enough for the contributions of the bound electrons to be considerable. Recent work, to be published shortly, shows that this is indeed the case, and that a metal containing free electrons only can have no X -point.

AUTHOR'S reply. I am very grateful to Dr. Price for the comments he has made in his discussion. It is true that one solution of my equation (15) is $x = a/2$. As Dr. Price himself points out, this gives rise to an X -point at $\lambda = 0.1 \mu$; this wavelength is so far in the ultra-violet region that my original assumption of free electrons being responsible for the mechanism of reflection cannot be maintained. Therefore this solution can be rejected. On dividing (15) by $2x - a$ there remains the quadratic $2x^2 - (3a - 2y)x + ay = 0$, which gives

$$X = \frac{1}{4}\{3a - 2y \pm \sqrt{(9a^2 - 20ay + 4y^2)}\} \approx \frac{1}{4}\left[3a - 2y \pm 3a\left(1 - \frac{10y}{9a} + \frac{2y^2}{9a^2} \dots\right)\right]. \dots (1)$$

The positive sign gives an X -point which falls still into the region where the effect of bound electrons predominates, and only the negative sign yields a value comparable with those obtained experimentally by Dr. Price. Since $y/a \approx 1/700$ it follows from (1) above that $x = \frac{1}{4}(-2y + 10y/3 \dots) = y/3$, which yields my equation (16). Since (16) has been tested for nickel by computing values for λ_r from katoptric data, and the λ_x predicted by this equation is confirmed experimentally, I fail to see why the conclusions drawn from this formula should be unsound. Nor can I understand how Dr. Price arrives at the conclusion that stationary points in $dR/d\nu$ presuppose negative values for the reflectivity. However, experiments are in progress which are designed to test the temperature variation of λ_x ; results will be reported on completion of this work.

DISCUSSION

on paper by G. F. J. GARLICK and A. F. GIBSON entitled "The Electron Trap Mechanism of Luminescence in Sulphide and Silicate Phosphors" (*Proc. Phys. Soc.*, 1948, **60**, 574).

Dr. H. A. KLASSENS. Lord, Rees and Wise (*Proc. Phys. Soc.*, 1947, **59**, 473), have explained their decay results also by assuming the electrons to be trapped near excited centres, but they allow for the possibility of such electrons escaping to the conduction band and combining with other excited centres. Does Dr. Garlick allow for this process? Also, how does he explain that Cu introduced into zinc sulphide activated with Mn prolongs the afterglow of the Mn emission?

AUTHORS' reply. The assumptions made by Lord, Rees and Wise in their paper are not necessary to explain our experimental results. They are based on very restricted phosphorescence decay measurements on one specimen of zinc sulphide activated by zinc and silver with a trace of copper. No independent study of the trapping states was made using thermoluminescence measurements. It would be difficult to distinguish the effects of certain traps being related to different emission centres in such a phosphor since the emission spectra of the two different centres formed by zinc and silver are almost identical.

With respect to the effect of copper in prolonging the afterglow of zinc sulphide phosphors activated by manganese, it must be assumed that electrons from traps associated with the copper impurity cause emission in manganese centres when they escape. However, it is not essential that the transit between centres should take place via the conduction levels of the matrix. Such a process, involving entry of electrons into the conduction levels, would be temperature dependent, whereas in some specimens of this type of phosphor we have found no marked temperature effect. Since electron traps have a large effective size (Garlick and Gibson, *Proc. Roy. Soc. A*, 1947, **188**, 485), "overlap" of trapping states belonging to different centres is likely which may allow electron transfers. This "overlap" of traps can also explain the difference in the copper concentrations for optimum afterglow and optimum emission efficiency in zinc sulphide phosphors.

REVIEWS OF BOOKS

Science and the Meanings of Truth, by MARTIN JOHNSON. Pp. 179. (London : Faber & Faber, 1946.) 12s. 6d. net.

This book continues the line of thought which Dr. Johnson has followed in two earlier works, the emphasis here being laid on the meanings of "truth" in modern physics. (Somewhat unfortunately the words "truth" and "truthfulness" are used indiscriminately: the reader who is accustomed to understanding by the latter the laudable but limited quality exhibited by the youthful George Washington must learn to give it a wider meaning here.) In the older physics causal explanations of phenomena in terms of inferred entities resembling phenomena were sought through hypotheses, and a hypothesis was true if it explained the facts without exception. In modern physics this aim is abandoned and it is recognized that physics reveals only a structure characterizing entities about which we know nothing. The word "truth" is thus left without a meaning, and the author suggests that it should be identified with "communicability" and "coherence" as characteristics of the structure. The arguments for this view occupy the major part of the book. The remainder is concerned with the philosophical problem of the relation between the structure and our sense experience and finally with the relation between the scientific and non-scientific realms of experience. For reasons of space, only the first part, which most directly affects physicists, will be discussed in this review.

The book reveals, as would be expected, an intimate knowledge of the present situation in physics and allied departments of thought. It is not easy reading but it repays the necessary effort. The fundamental problems aroused by relativity and quantum theory are considered in relation to the modern branch of logic known as "logistics", especially the calculus of propositional functions, but strangely enough the post-logistics developments, such as those of the metamathematicians and the intuitionists, which transcend logistics almost as thoroughly as that transcended scholastic logic and are much more vital for the subject treated, are not mentioned. On the question whether the laws of nature are empirical or rational in origin, Dr. Johnson's view seems to be that they are both. As a disciple of Milne he thinks that they can be derived from "our intuitive appreciation of the passage of time", which is taken as empirical, without further observation, by the use of reason alone. At least, I think that this is his view.

The core of the book, however, is the thesis that "communicability" is the dominating criterion of truth in physics. Dr. Johnson has emphasized the importance of communicability in each of his previous works, and it is evidently invested in his mind with profound significance. Unfortunately, however, he uses the word in so many senses that it ceases to

have any consistent meaning. In one place a thing is communicable if it can be communicated from one observer to another. That is legitimate, but shortly afterwards it is implied that the two observers must *agree* about what is communicated, and that is not at all implied in the proper use of the word. Yet again, a scientific proposition, to be true, must be "communicable from one experimental situation to another", by which is meant that a conception, to be true, must be applicable to different, originally independent, fields of investigation. For example, whenever a wave-like property, such as a measurable frequency or a diffraction pattern, appears in a communicable set of experiments, then this particular set achieves communication in turn with other situations which have been characterized by wave-like properties. This state of affairs is also called "coherence". If this variation of meaning merely caused confusion it would be a remediable nuisance, but it is in fact a fatal flaw because a situation, having been endowed with communicability in one sense, is thereupon assumed to possess it in others, and so a fallacy is introduced. It is easy to prove that black is white if one is allowed to call them both grey.

The most striking instance occurs in the treatment of the Lorentz transformation. It is a fact that this transformation makes different observations (such as those of the orbital motion of the Earth and the non-displacement of fringes in the Michelson-Morley experiment) coherent. Hence it establishes "communicability" and so is assumed to demand the existence of a host of observers to make the communications which enable them to agree with one another. All this is simply false. The Lorentz transformation has nothing to do with different observers; it concerns the way in which a single observer changes his coordinate system. This is not opinion, it is obvious fact. Lorentz did not derive the transformation so that he could agree with the Man in the Moon. He derived it so that different observations by what was effectively one observer—an earthbound one—could be reconciled. Hence, since virtually only one observer is concerned, the question of "communicability" in the proper sense of the word does not arise. The position can perhaps be seen more clearly in relation to a simpler change of coordinate systems—one in which we merely change the origin of coordinates. The statement that two places are respectively 10° and 15° east of Greenwich is rejected as a fundamental statement not because it is not communicable; it is communicable and, furthermore, both parties to the communication can agree about it, and the statement preserves the coherence of all geography. But it is rejected because Greenwich is an arbitrary starting-point and the simpler statement that the places differ in longitude by 5° is preferred because it is free from such arbitrariness. That has nothing to do with different observers; it holds good if there is only one observer in the world. With the Lorentz transformation the position is exactly the same, with velocities substituted for positions. That also has nothing to do with a multitude of observers; it is concerned with the various coordinate systems that a single observer may choose. The illusion that different observers, and communicability, are involved is the result of the most unfortunate device that early expositors adopted in order to explain the theory of relativity to the man in the street. He was assumed not to be familiar with coordinate systems but to have a natural tendency to think himself at rest (no man of science has done so since shortly after the time of Copernicus). The second coordinate system was therefore personified into a second observer who also thought himself at rest.

All this, I repeat, is not opinion but simple, undeniable fact. As in his earlier works, Dr. Johnson overlooks it and bases his argument on the supposed actuality of imaginary observers. He is not a dishonest thinker, and one can only conclude that he is so charmed by Milne's theory of kinematical relativity, his devotion to which is manifest throughout his writings, that he cannot bring himself to discard the fictitious situation on which Milne himself bases it. For if—acknowledging as he does, the Lorentz transformation to be "the classic instance" of what he calls the "attainment of communicability"—he takes that transformation in its legitimate, unperverted form, he will find that he cannot proceed to the conclusion he so much desires, and the whole theory of kinematical relativity will be exposed as a baseless structure. "Communicability" makes a host of observers a necessity for the *possibility of existence*, not merely the practicability of prosecution, of science, and this is fundamentally false. A single mind, with unlimited intelligence, mechanical skill and freedom of movement, and sufficient time, could derive and verify the whole of our physical knowledge without anything with which it could "communicate". If any one doubts this he has only to try to name a piece of knowledge not so obtainable. Any philosophy which is inconsistent with this fact stands self-condemned.

What makes this characteristic of Dr. Johnson's work the more regrettable is that he is one of the few writers who, in this age of specialization, have the breadth of vision and the depth of thought necessary to grasp the meaning of the scientific movement as a whole and estimate its significance. He has something to say and has the ability to say it. It is profoundly to be hoped that he will have the courage to discard fanciful speculations and reform his contribution to the philosophy of our time so that it can rest on foundations true in the George Washington sense of the word.

HERBERT DINGLE.

Magnetic Materials, by F. BRAILSFORD. Pp. ix+156. (London: Methuen & Co. Ltd., 1948.) 6s.

Dr. Brailsford, as a member of the Research Department of the Metropolitan-Vickers Electrical Company, has had ample opportunity of first-hand acquaintance with the many new magnetic materials which have found, and are still finding, their way into all kinds of laboratories and workshops. It is natural that he should open his survey of these materials and their properties with a short, but adequate, introduction to modern theories of ferromagnetism, the properties of ferromagnetic single crystals and the various factors which determine the magnetic behaviour of metals and their alloys. This introduction, which takes up more than half of the booklet, is excellently illustrated with clear diagrams.

The materials themselves are studied under three headings: Iron and Silicon-Iron Alloys, Nickel-Iron and other Alloys, and Permanent Magnet Materials. All students should find these sections very helpful. Lists of original references are given, and it should therefore be easy to follow up particular points of interest. The author has certainly packed a great deal of useful information into a very small space, mainly by the wise use of graphs. The book can be unreservedly recommended to all interested in the subject.

The reviewer was struck by the use of the term *retentivity* on page 135 to denote *residual inductance* or *remanence*, for which Ewing used the term *retentiveness*. As far as can be ascertained, the word *retentivity* has hitherto been used to denote residual intensity of magnetization.

L. F. BATES.

Précis de Physique Générale, II: La Chaleur, by P. and A. MERCIER. Pp. 105. (Neuchâtel: Editions du Griffon, 1947.) 12 fr. (Swiss).

One of the authors is an engineer, former dean of the scientific faculty at the Collège de Genève, and the other is professor of theoretical physics at Berne. Their joint work should, therefore, present a good balance between the theoretical and the practical outlooks. They have, in fact, written a very good summary of the fundamentals of classical heat theory. Classical thermodynamics is presented in a very sound manner, with a very clear treatment of the basis of the second law and the definition of absolute temperature, but there is no reference to modern quantum mechanics and none to statistical mechanics. Of the subjects dealt with, it is interesting to note that whilst internal energy and entropy are introduced, the only functions derived from them is Gibbs' function, here called the *thermodynamic potential*. No mention is made of the function known as heat content or enthalpy, nor of Helmholtz's free energy.

The book is thus very good for those who desire a brief development of thermodynamics, with the practical side of thermometry and calorimetry, but it needs supplementing for such matters as heat conduction and radiation.

J. H. A.

THE PROCEEDINGS OF THE PHYSICAL SOCIETY

VOL. 61, PART 2

1 August 1948

No. 344

Shock Waves in Air at Very High Pressures *

By D. R. DAVIES

Communicated by W. G. Penney ; MS. received 5 September 1947, and in amended form 18 February 1948 ; read 9 January 1948

ABSTRACT. The equations of the shock wave in air at N.T.P. are solved numerically for pressures on the high pressure side up to 1000 atmospheres. The solution enables all the physical entities, such as temperature, wave velocity, particle velocity, air density, etc., on the shock wave front to be expressed numerically as a function of the wave pressure. To solve these equations it is necessary to know the internal energy (E) and volume (v) of one gramme of air over certain regions of the two-dimensional range $1 < p < 1000$ atmospheres and $273 < T < 16000^\circ \text{K}$. Calculations are carried out to assess the numbers of the various types of molecules, atoms and ions present at any p and T . The only E values needed at high p are those for which T is also high, and the simple gas laws may therefore be assumed for the volume determinations. The resultant energy is obtained by summing the internal energies of the different groups of atoms and molecules present. Statistical mechanics furnishes equations whose solution fixes the composition. This set of equations is difficult to solve to any degree of accuracy if oxides of nitrogen are taken into account ; of these only nitric oxide, NO, ever exceeds a concentration of 1% by weight, and its maximum concentration is less than 5%. As a first approximation, the shock wave equations are worked out on the assumption that the species present are N_2 , O_2 , N and O. These calculations make full allowance for all quantum states apart from ionic states. They are refined later by taking into consideration the presence of NO and argon (1.3% by weight) and ionization possibilities.

§ 1. INTRODUCTION

DURING the early part of the recent war the problem of working out theoretically the distribution of pressure and velocity *behind* a spherical shock wave in water was resolved by solving the hydrodynamical equations for the motion of water behind the waves by numerical methods. The similar problem of propagation of shock waves in *air*, however, proved to be much more difficult, as the equation of state of air at high temperatures presented considerable difficulties. The first step in the solution consisted of working out the relationship between shock wave velocity, mass velocity, temperature, pressure, etc., on the shock wave itself, assuming a sudden discontinuity at the wave front. It is this problem which is described in the paper.

A system of equations was derived by Lewis and von Elbe (1938) relating the various physical entities connected with a shock wave front. The notation used in the paper for these physical entities is as follows : shock wave velocity U , mass velocity u , temperature T , volume of unit mass of air v , internal energy per unit mass of air E , and the air pressure p . These shock wave equations may be solved if we have a complete knowledge of the *internal energy of air* in the neighbourhood of the shock wave conditions. They completely determine

* This paper forms part of a thesis presented to the University of London for the Ph.D. degree.

all the variables U , u , T , p , v , etc., if one of the set is known, in terms of the known conditions ahead of the wave front. It is assumed in this paper that the air on the low pressure side is at N.T.P. The independent variable chosen is p , the pressure on the high pressure side. Provided that only differences of pressure amounting to not more than a few atmospheres on the two sides of the wave are considered, simple calculations lead to reliable values of U , u , and T , etc., in terms of p , as worked out by Lewis and von Elbe.

Solution becomes difficult for p values above five atmospheres due to the variation of the specific heats of air from the values obtained at ordinary temperatures and pressures, as very high temperatures are generated on the wave front at high pressures. For pressures up to about thirty atmospheres the only complication to the simple theory arises out of the necessity to allow for the vibrational energy of air molecules. This paper is concerned with the pressure range from thirty to one thousand atmospheres, where, due to several complications, an elaborate calculation becomes necessary. Dissociation of the air molecules must be taken into account: molecules and atoms appear on electronic energy levels other than their ground states, and allowances must be made, too, for the changing composition of the air as the pressure and temperature conditions change. At pressures greater than one thousand atmospheres ionization complicates matters further, although dissociation has ceased to be troublesome. These calculations are not treated in this paper.

It becomes necessary as a first step to evaluate the internal energy E and the volume v of one gramme of air over most of the range $1 < p < 1000$ atmospheres and $273 < T < 16\,000^\circ \text{K}$. At this first stage in the solution, standard methods of statistical mechanics furnish equations whose solution fixes the composition of the air. If great accuracy is required, these equations provide numerical difficulties. In addition to the formation of nitrogen and oxygen atoms, argon and oxides of nitrogen have to be dealt with. However, only nitric oxide, NO, and argon, A, ever appear in concentrations greater than 1% by weight. Once the composition has been fixed, the volume follows immediately from the law $p = nkt$; and the internal energy is the sum of the internal energies of the different groups of atoms and molecules present.

As a first approximation only, nitrogen and oxygen molecules and atoms are considered, and E and v are evaluated over the range specified above. These functions enable the shock wave equations to be solved and thus specify approximately the neighbourhood within which the wave conditions lie. These calculations are later refined by taking into account (a) formation of NO, (b) formation of argon and (c) a more recent evaluation by A. G. Gaydon and W. G. Penney (1945) of the dissociation energy of nitrogen. This later work is greatly simplified, as the results of the first approximation considerably narrow the two-dimensional range of T and p , within which the numerical work has to be carried out.

§2. INTERNAL ENERGY PER GRAMME FOR THE MAIN CONSTITUENTS OF AIR UP TO $16\,000^\circ \text{K}$.

Air at normal temperatures and pressures consists primarily of oxygen and nitrogen molecules with a small proportion of inert gases, mainly argon. The percentages by weight are: nitrogen 75.5, oxygen 23.2, argon 1.3. At a temperature of about 2000°K . the oxygen commences to dissociate, and near 3000°K . the nitrogen constituent behaves similarly. At the 4000°K . stage a considerable

proportion of the nitrogen atoms combine with oxygen atoms to form oxides of nitrogen. Of these only nitric oxide ever exceeds the figure of 1% by weight. The components considered in this calculation are nitrogen and oxygen atoms and molecules, nitric oxide and argon. The proportion by weight of these gases, in one gramme of air, over the temperature range 4000–16000° K. varies considerably with temperature and pressure conditions. It becomes necessary, therefore, at the outset, to evaluate the energy per gramme of these components over the temperature range and later to assess the degrees of dissociation at each temperature.

Statistical thermodynamic theory leads to simple formulae for the translational, rotational, vibrational and electronic energies of the various components. These are: (i) the internal energy per molecule, or atom, associated with the three-dimensional degrees of freedom

$$E_t = 3kT/2, \quad \dots\dots(1)$$

where k = Boltzmann's constant; (ii) the internal energy per molecule of the two rotational degrees of freedom of a diatomic molecule,

$$E_r = kT; \quad \dots\dots(2)$$

(iii) the internal energy of vibration of a molecule, per molecule,

$$E_v = h\nu / [\exp(h\nu/kT) - 1], \quad \dots\dots(3)$$

where ν is the fundamental vibration frequency and h is Planck's constant; (iv) the internal energy per molecule or atom, associated with the electronic motion

$$E_e = \{ \sum_i g_i W_i \exp(-W_i/kT) \} / \{ \sum_i g_i \exp(-W_i/kT) \}, \quad \dots\dots(4)$$

where g_i is the statistical weight and W_i the energy of the electronic level i . The ground state of an atom is taken to be $D/2$, where D is the energy of dissociation of the ground state of the molecule formed from two of the atoms.

These formulae, and later ones covering the equilibrium constants which specify the degrees of dissociation, are based on certain necessary approximations. It is assumed that complete factorization of the partition functions, §3, upon which the energy functions and dissociation formulae are based, is exact, hence that the total energy is the sum of the separate energies. This is not strictly true. The spacing of the vibrational levels is not constant, and the moment of inertia of the molecule changes with increasing vibration and from one electronic level to another. However, the errors introduced by these approximations are not large, especially in view of the order of accuracy required in the solution of the shock wave equations.

Table 1 gives the spectroscopic data used in the energy and dissociation formulae.

Table 1

	D (ev.)	ν (cm ⁻¹)	B_e (cm ⁻¹)
O ₂	5.082	1580	1.440
N ₂	9.764	2360	1.998
NO	6.490	1900	1.684

The relation between B_e and I , the moment of inertia, is $B_e = h/(8\pi^2 Ic)$, where c is the velocity of light (3×10^{10} cm/sec.).

Other spectroscopic data that we require may be summarized. The ground state of O is $^3\Sigma_1$; there is another state $^1\Delta$, at 0.96 ev., and a second excited state $^1\Sigma$, at 1.62 ev. The ground state of O is ^3P ; there are excited ^1D at 1.97 ev. and ^1S at 4.5 ev. The ground state of N_2 is $^1\Sigma$ and the ground state of N is ^4S . The excited states of N_2 and N are too high to be considered as they have very small Boltzmann factors. These $^1\Sigma$ and ^1S states of O_2 and O have little effect, but they have been included. The ground state of NO is $^2\Pi$ and the first excited state $^2\Sigma$. High excited levels have no significant influence on the energy functions.

The statistical weights of the excited states may be written down once the electronic levels are known. For example, statistical weight of the ground state of O is 9, because the level has spin degeneracy 3 (triplet nature) and orbital degeneracy 3 (P states). The multiplet structure has been neglected because the overall separation of the ^3P multiplet is somewhat less than kT at temperatures producing appreciable dissociation of the oxygen molecules.

Statistical weights used for the ground state, first and second excited levels respectively are : for O_2 —3, 5, 1, for O—9, 5, 1 and for NO—2 and 2 for the ground state and first excited level.

Table 2. Internal energy of air constituents

Substance	T ($^{\circ}$ K.)	$E_t + E_r + E_v$ (cal/gm.mol.)	Electronic energy (cal/gm.mol.)	Total internal energy (cal/gm.)
Oxygen molecules	2,000	12,060	142	381
	6,000	39,600	4,975	1,393
	10,000	67,330	8,800	2,379
Oxygen atoms	4,000	11,910	86	750
	8,000	23,830	1,454	1,580
	12,000	35,750	3,620	2,460
	16,000	47,660	5,722	3,337
Nitrogen molecules	2,000	11,450	0	409
	6,000	38,710	0	1,382
	10,000	66,300	0	2,368
	14,000	94,124	0	3,362
Nitrogen atoms	2,000	5,960	0	426
	6,000	17,870	0	1,277
	10,000	29,790	0	2,128
	14,000	41,707	0	2,978
Nitric oxide	2,000	11,794	0	393
	6,000	39,150	4	1,305
	10,000	66,930	227	2,238
	14,000	94,696	1,366	3,202
Argon	2,000	5,957	0	149
	6,000	17,874	0	448
	10,000	29,790	0	746
	14,000	41,707	0	1,193

Note : Intermediate values for this and ensuing tables may be obtained on application to the author.

The formulae used for the electronic energies in cal/gm. atom or molecule : for oxygen molecules,

$$E_{el} = 23.07 \times 10^3 (4.8 e^{-x_1} + 1.62 e^{-x_2}) / (3 + 5 e^{-x_1} + e^{-x_2}); \dots (5)$$

for oxygen atoms,

$$E_{\text{el}} = 23.07 \times 10^3 (9.835 e^{-v_1} + 4.5 e^{-v_2}) / (9 + 5e^{-v_1} + e^{-v_2}); \dots\dots (6)$$

for nitric oxide molecules,

$$E_{\text{el}} = 23.07 \times 10^3 (5.453 e^{-z_1}) / (1 + e^{-z_1}), \dots\dots (7)$$

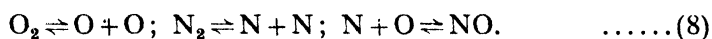
where $x_1 = P_1/kT$, $x_2 = P_2/kT$, $y_1 = Q_1/kT$, $y_2 = Q_2/kT$, $z_1 = R_1/kT$, and P_1 , Q_1 , R_1 are the energies of the first excited levels and P_2 , Q_2 the energies of the second excited levels, all in appropriate units. The relationship $23.07 \text{ cal/gm. atom} = 1 \text{ e.v.}$ is used.

In table 2 the energies per gramme of each constituent of air are given as a function of the temperature.

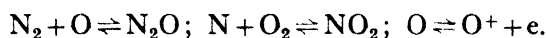
§ 3. ESTIMATION OF THE EQUILIBRIUM CONSTANTS OF NITROGEN, OXYGEN AND NITRIC OXIDE UP TO 16000° K.

It is possible with the aid of the general laws of thermodynamics to predict the form of the functional dependence of the degree of dissociation of a gas on the temperature and pressure conditions. These laws, however, leave unspecified certain constants. A knowledge of the energy levels which the particles of the gas can occupy, as furnished by standard spectroscopic methods, enables these constants to be evaluated; the degree of dissociation can thus be computed numerically from thermodynamics in conjunction with spectroscopic data.

The chemical reactions to be considered are:—



Many more, whose effect is small and have been neglected, may occur, such as



Consider the simplest type of dissociation equilibrium, $\text{AB} \rightleftharpoons \text{A} + \text{B}$, in which gaseous compound AB dissociates into the monatomic vapours A and B. Then the equilibrium constant K is defined by $K = p_A \cdot p_B / p_{\text{AB}}$, where p_A and p_B are the partial pressures of the monatomic vapours and p_{AB} the partial pressure of the original compound.

Using this definition in the nitrogen reaction $\text{N}_2 \rightleftharpoons 2\text{N}$, it is easily shown that

$$a_1 = \sqrt{(K_1 / (K_1 + 4p_1))}, \dots\dots (9)$$

or $\ln K_1 = \ln[4a_1^2 p_1 / (1 - a_1^2)]$, where a_1 is the ratio = (weight of nitrogen atoms in a given volume) / (weight of nitrogen atoms plus molecules in the same volume), p_1 is the partial pressure of nitrogen ($\text{N} + \text{N}_2$), and K_1 the equilibrium constant of the reaction at temperature T .

A similar equation holds for oxygen (a_2). Hence, once the K value is fixed at a given temperature, the dissociation, given by a_1 etc., can be evaluated.

Calculation of equilibrium constant as a function of temperature is simple once the partition functions for the components have been obtained.

The partition function P_1 for O and N may both be expressed as a product of three linear translational partition functions and one electronic partition function $P_1 = T^3(m)P(A)$. A represents the atoms O or N, m the mass of the atom.

Similarly the P_2 functions for nitrogen and oxygen molecules may be written $P_2 = T^3(2m) \cdot R^2(I) \cdot V(\nu) \cdot P(M)/2$, where T represents translation, R represents

end-over-end rotations, V represents vibration, and $P(M)$ represents the electronic part of the partition function of a molecule M . The factor $\frac{1}{2}$ enters for N_2 and O_2 , because only one half of the rotational phase space is available to these homonuclear molecules N_2 and O_2 . It does not enter for NO .

These standard expressions are used:

$$\begin{aligned} T(m) &= \nu^{\frac{1}{2}} \sqrt{(2\pi m k T)}; \quad R(I) = 2\pi \sqrt{(2\pi I k T)}; \\ P(O) &= 9 + 5 \exp(-Q_1/kT) + \exp(-Q_2/kT) + \dots; \\ P(O_2) &= 3 + 5 \exp(-P_1/kT) + \exp(-P_2/kT) + \dots; \\ P(NO) &= 2 + 2 \exp(-R_1/kT) + \dots; \\ P(N) &= 4; \quad P(N_2) = 1; \quad V(\nu) = 1/[1 - \exp(-h\nu/kT)]. \end{aligned}$$

For the reaction $AB \rightleftharpoons A + B$, the equation obtained for $\ln K$ is

$$\begin{aligned} \ln K &= -D/kT + 3/2 \ln[2\pi M_A M_B kT/h^2 M_{AB}] + \ln h B_e + \ln\{1 - \exp(-h\nu/kT)\} \\ &+ \ln g_0^A g_0^B / g_0^{AB} + \ln \frac{[1 + (g_1^A/g_0^A) \exp(-X/kT)][1 + (g_1^B/g_0^B) \exp(-Y/kT)]}{[1 + (g_1^{AB}/g_0^{AB}) \exp(-Z/kT)]}, \end{aligned} \quad \dots\dots(10)$$

where D = energy of dissociation from the lowest vibrational level; M_A, M_B = mass of atom A, B; $M_{AB} = (M_A + M_B)$ = mass of molecule AB; g_0, g_1 = statistical weight of atoms or molecules in the ground state and first level (the superscripts are self-explanatory); X, Y, Z = energy increase from ground state to the first level for atoms A, B, molecules AB respectively.

The second and higher electronic levels have no effect. On the basis of wave mechanics there must for N_2 and O_2 be added, to the right hand of the above equation, a term $\ln 2$ to allow for the symmetry of molecules AA.

The expression used for the oxygen reaction is

$$\begin{aligned} \ln K &= -D_{O_2}/kT + 3/2 \ln \pi m_O kT/h^2 + \ln h B_e + \ln[1 - \exp(h\nu_{O_2}/kT)] + \ln 54 \\ &+ 2 \ln[1 + 0.556 \exp(-Q_1/kT)] - \ln[1 + 1.667 \exp(-P_1/kT)], \end{aligned} \quad \dots\dots(11)$$

(m_O = mass of oxygen atom). Similarly for nitrogen

$$\ln K = -D_N/kT + 3/2 \ln \pi m_N kT/h^2 + \ln h B_e + \ln[1 - \exp(-h\nu_N/kT)] + \ln 32 \quad \dots\dots(12)$$

(m_N = mass of nitrogen atoms), and for nitric oxide

$$\begin{aligned} \ln K &= -D_{NO}/kT + 3/2 \ln 2\pi m_N m_O kT/h^2 m_{NO} + \ln h B_e + \ln[1 - \exp(-h\nu_{NO}/kT)] \\ &+ \ln 18 + \ln[1 + 0.556 \exp(-Q_1/kT)] - \ln[1 + \exp\{-R_1/kT\}]. \end{aligned} \quad \dots\dots(13)$$

The B_e in this formula is c times that given in table 1, since frequencies are usually expressed in wave numbers as in table 1.

Table 3. Equilibrium constants of oxygen, nitrogen and nitric oxide as functions of temperature

T ($^{\circ}$ K.)	K (O) (dyne/cm ²)	K (N) (dyne/cm ²)	K (NO) (dyne/cm ²)
2,000	0.595	0	0
4,000	3.509×10^6	3.390	2.754×10^4
8,000	6.740×10^9	7.959×10^6	8.715×10^8
12,000	9.488×10^{10}	1.158×10^9	3.623×10^{10}
16,000	3.929×10^{11}	1.458×10^{10}	2.620×10^{11}

§ 4. ESTIMATION OF INTERNAL ENERGY $E(p, T)$, AND VOLUME, $v(p, T)$, FOR ONE GRAMME OF AIR

In this section the effects of oxides of nitrogen and argon are neglected and a D value of 7.35 ev. is assumed for nitrogen. The general formulae used are

$$E = \sum_i n_i(p, T) \cdot E_i(T) \quad \dots\dots (14)$$

and $\sum_i n_i(p, T) = n(p, T)$, where $n_i(p, T)$ is the number of any class of atoms or molecules i , $E_i(T)$ is the internal energy per particle of this type of particle, E the total internal energy per gramme of air, and

$$pv = n(p, T)kT. \quad \dots\dots (15)$$

Fortunately the only values of E required at high pressures are those for which the temperature is also high. At no point does v become less than about 1/60 of the volume at N.T.P. so that the use of equation (15) is justified.

By means of equations (9), (11) and (12), a_1 and a_2 may be evaluated over the dissociation region, and hence by (14) the energies per gramme of nitrogen and per gramme of oxygen as functions of p_1 or p_2 and T . It can be shown that

$$(1 + a_1)/p_1 = 0.269(1 + a_2)/p_2, \quad \dots\dots (16)$$

where p_1, p_2 are the partial pressures of nitrogen and oxygen respectively. Hence p_1 and p_2 may be evaluated at each temperature for the required value of total pressure $p (= p_1 + p_2)$. Interpolating at these p_1 and p_2 values in the energy curves for oxygen and nitrogen, and assuming that one gramme of air always contains 24% by weight of oxygen and 76% by weight of nitrogen, we obtain the final curves for the internal energy per gramme of air for $273 < T < 16000^\circ \text{K}$. and total pressures equal to 15, 25, 35, 45, 65, 75, 150, 200, 400, 600, 800 and 1000 atmospheres; thus later interpolation between two p curves is easily performed.

The general character of these results is seen in figure 1. Once the p_1 and p_2 values have been determined for a given p , the volume of one gramme of air is given by

$$v = \frac{\frac{1}{2}kT}{(p_1 m_N / (1 + a_1)) + (p_2 m_O / (1 + a_2))}. \quad \dots\dots (17)$$

Some of these values are indicated in figures 1 and 2.

It must be borne in mind, however, that these curves are intended to give an indication only of the actual trend of the E and v curves. They depend on a previous determination of D for nitrogen and neglect the formation of oxides of nitrogen and the presence of inert gases, etc.

§ 5. SOLUTION OF THE SHOCK WAVE EQUATIONS: FIRST APPROXIMATION

Consider a coordinate system moving with the wave front. The mechanism of transfer across the wave front is stationary, and consequently the mass, momentum and energy of a gas passing across the wave front must be conserved. According to Lewis and von Elbe (1938),

$$u_1/v_1 = u_2/v_2; \quad p_1 + u_1^2/v_1 = p_2 + u_2^2/v_2; \quad E_1 + u_1^2/2 + p_1 v_1 = E_2 + u_2^2/2 + p_2 v_2.$$

From these it follows that

$$\Delta E = \frac{1}{2}(p_2 + p_1)(v_1 - v_2), \quad \dots\dots (18)$$

$$U = u_1 = v_1 \sqrt{\{(p_2 - p_1)/(v_1 - v_2)\}}, \quad \dots\dots (19)$$

$$u = u_1 - u_2 = \sqrt{\{(p_2 - p_1)(v_1 - v_2)\}}, \quad \dots\dots (20)$$

where p_1, v_1 refer to the low-pressure side and p_2, v_2 to the high-pressure side.

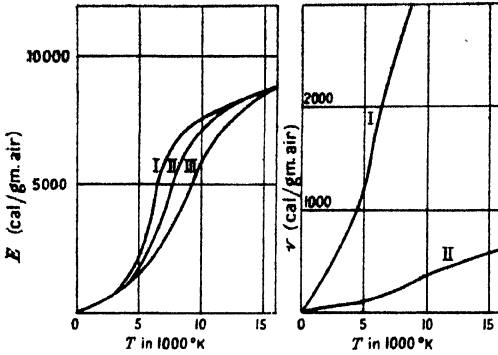


Figure 1. Internal energy of air at constant pressure for varying T , neglecting the presence of NO and A. $D_0 = 7.35$ ev.; I, $p = 15$ atmos.; II, $p = 150$ atmos.; III, $p = 1000$ atmos.

Figure 2. Volume of air at constant pressure for varying T , on assumptions of figure 1.

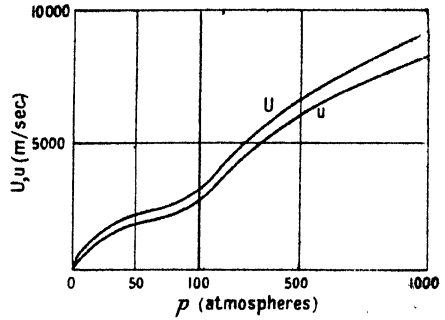


Figure 3. Shock wave velocity U and mass velocity u as a function of p . (Note: two scales for p .)

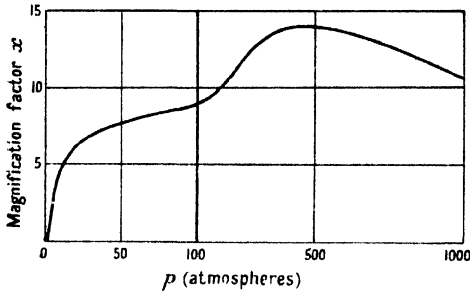


Figure 4. Magnification factor x . (Note: two scales for p .)

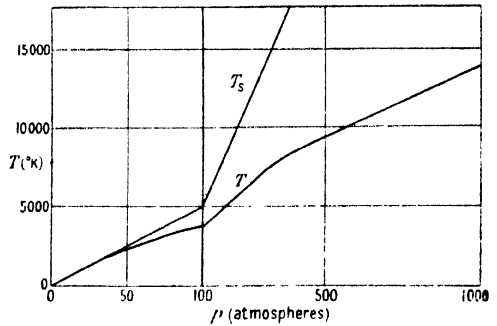


Figure 5. The temperature T behind the shock wave. The temperature T_s is also shown, where T_s is calculated on the assumption that $\gamma = 1.4$ for all T and p . Two scales are used.

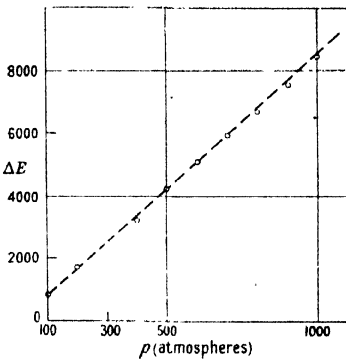


Figure 6. The change ΔE in the internal energy per gramme of air, due to the passage of a shock wave, as a function of the pressure on the wave. ΔE is in calories per gramme of air; p in atmospheres.

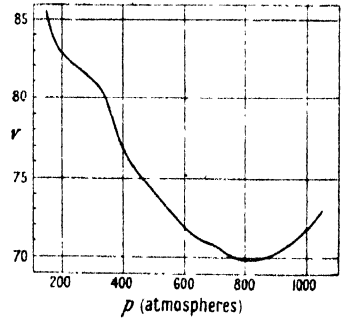


Figure 7. The volume (in cm^3) of one gramme of air as a function of the shock wave pressure (in atmos.).

These equations may be solved numerically by a process of successive approximation, starting with p_2 and assessing a value for v_2 . The volume and energy curves then give T_2 and E_2 . If on one side the air is at N.T.P., E_1 is known, and hence ΔE . By comparison with the value of ΔE calculated from (18), adjustment may be made to v_2 . Values of ΔE , T_2 , U and u follow directly. These are given in table 4.

Table 4. Solution of the shock wave equations (neglecting oxides of nitrogen and argon; and assuming D for nitrogen to be 7.35 ev.)

p_2 (atmospheres)	30	100	500	1,000
U (metres/sec.)	1,670	3,000	6,560	9,350
u (metres/sec.)	1,370	2,590	5,960	8,310
T_2 ($^{\circ}$ K.)	1,480	3,600	7,950	15,000
E_2 (calories per gm. of air)	287	875	4,320	8,450
v_2 (cm ³ per gm. of air)	140	107	70	86

§6. SOLUTION OF THE SHOCK WAVE EQUATIONS: INCORPORATING EFFECT OF NO AND A

In this solution the value of D for nitrogen is 9.764.

It has been shown that the effect of allowing for the formation of NO is a small one, resulting at its maximum (4000–5000 $^{\circ}$ K.) in an increase in the value of E_2 of about 12 calories and a decrease of p_2 of about 5 atmospheres. Allowance for argon has a similarly small effect. The increase in the nitrogen dissociation value has a much larger influence on the shock wave conditions. The first solution is unchanged up to about 100 atmospheres.

The dissociation equations are easily written down. Consider the species O_2 , O, N_2 , N, NO and argon. The last component does not change, so that its influence is easily predicted. If v is the volume of one gramme of air at temperature T , α the number of oxygen atoms, present as O, O_2 or in NO, in one gramme of air, β the number of nitrogen atoms, present as N, N_2 or in NO, in one gramme of air, x , y the fraction of oxygen, nitrogen, in the form of atoms, and z is the number of molecules of NO, the equations relating x , y , z , v and p are

$$x^2 = [1 - x - z] v K_1 / 2\alpha kT, \quad \dots\dots (21)$$

$$y^2 = [1 - y - z\alpha/\beta] v K_2 / 2\beta kT, \quad \dots\dots (22)$$

$$z = (xy/v)\beta kT / K_3, \quad \dots\dots (23)$$

$$p' = (kT/2v)[\alpha(x+1) + y(\beta+1)], \quad \dots\dots (24)$$

$$p = p' + p_A, \quad \dots\dots (25)$$

where K_1 , K_2 , K_3 are the equilibrium constants at temperature T for oxygen, nitrogen and nitric oxide respectively, p_A = partial pressure of the argon and p' = partial pressure of remaining constituents.

From (21), (22) and (23),

$$x = \frac{(1-z)(1-\alpha z/\beta)K_2/2K_3 - 2\alpha K_3 z^2/\beta K_1}{(1-\alpha z/\beta)K_2/2K_3 + \alpha K_2 z/\beta K_1}, \quad \dots\dots (26)$$

$$\text{and } y = \frac{(1-z)(1-\alpha z/\beta)K_2/2K_3 - 2\alpha K_3 z^2/\beta K_1}{(1-z)K_2/2K_3 + z}. \quad \dots\dots (27)$$

Substitution in (24) gives a bi-quadratic for z :

$$Az^4 + Bz^3 + Cz^2 + Dz + E = 0, \quad \dots\dots(28)$$

where

$$\begin{aligned} A = & + \left(\frac{\alpha}{\beta}\right)^2 \left[4 \left(\frac{K_3}{K_1}\right)^2 - 2 \left(\frac{K_2}{K_1}\right) + \frac{1}{4} \left(\frac{K_2}{K_3}\right)^2 \right] \\ & - \left(\frac{\alpha}{\beta}\right)^2 \left(\frac{K_3}{2p'}\right) \left(\frac{K_2}{2K_3} - \frac{2K_3}{K_1}\right) \left(1 - \frac{K_2}{K_3} + \frac{K_2}{K_1}\right); \\ B = & + \left(\frac{\alpha}{\beta}\right) \left(1 + \frac{\alpha}{\beta}\right) \left\{ \frac{2K_2}{K_1} - \frac{1}{2} \left(\frac{K_2}{K_3}\right)^2 \right\} - \left(\frac{\alpha}{\beta}\right) \left(\frac{K_3}{2p'}\right) \left[\left(\frac{\alpha}{\beta} + 1\right) \left(\frac{K_2}{K_3}\right) \left(\frac{K_2}{2K_3} - \frac{2K_3}{K_1}\right) \right] \\ & + \left(\frac{\alpha}{\beta}\right) \left(\frac{K_3}{2p'}\right) \left[\left(1 - \frac{K_2}{2K_3}\right) \left(\frac{\alpha K_2}{\beta K_1} - \frac{\alpha K_2}{\beta K_3} - \frac{K_2}{2K_3}\right) \right. \\ & \left. + \left(1 - \frac{K_2}{K_3} - \frac{\alpha K_2}{2\beta K_3}\right) \left(\frac{K_2}{K_1} - \frac{K_2}{2K_3}\right) \right]; \\ C = & + \frac{1}{2} \left(\frac{K_2}{K_3}\right)^2 \left[\frac{1}{2} \left(1 + \frac{\alpha}{\beta}\right)^2 + \frac{\alpha}{\beta} \right] - \left(\frac{2\alpha K_2}{\beta K_1}\right) \left(\frac{K_3}{2p'}\right) \left\{ \frac{\alpha K_2}{\beta K_3} \left[1 - \frac{K_2}{2K_3} \left(\frac{3}{2} + \frac{\alpha}{\beta}\right) + \frac{\alpha K_2}{2\beta K_1} \right] \right. \\ & \left. + \frac{K_2}{K_3} \left[\frac{\alpha}{\beta} \left(\frac{K_2}{K_1} - \frac{K_2}{2K_3}\right) + \frac{1}{2} \left(1 - \frac{K_2}{K_3} - \frac{\alpha K_2}{2\beta K_3}\right) \right] \right\}; \\ D = & - \frac{1}{2} \left(1 + \frac{\alpha}{\beta}\right) \left(\frac{K_2}{K_3}\right)^2 \left(1 + \frac{K_3}{2p'}\right); \\ E = & + \frac{1}{4} \left(\frac{K_2}{K_3}\right)^2. \end{aligned}$$

For a given T and p' the biquadratic can be solved for z ; x , y and v follow from equations (26), (27), and (23). From T and v , p_A follows, and hence p from (25). The x , y , z factors fix the composition of air and hence the component energies. The E_2 value is then the sum of these. This method is a very laborious one as it entails initially the evaluation of the coefficients A , B , etc., which in itself is not easy, and then the solution of the biquadratic for a large number of p' points.

An alternative procedure was developed which proved much simpler. Fixing the temperature T_2 , a value of z is chosen which, with the aid of (26), (27), (28), (24) and (25), yields a p_2 value in the neighbourhood of the shock wave conditions. It is here that the first solution is of use. With a little practice, choice of a z value at any T_2 to give an approximate p_2 becomes an easy matter. E_2 follows from x , y , z and hence ΔE . ΔE is also computed from equation (18). The correct "shock wave z " is the one which makes E equal by the two methods. The correct U , u , p_2 , v_2 , etc., for the given T_2 then follow from the shock wave equations.

At 3000° K. the nitrogen component has not commenced to dissociate appreciably, so that no nitric oxide has been formed at this stage. At 4000° K. air at shock wave conditions contains about 4.4%, at 5000° K. about 4.8% by weight of NO. The percentage then slowly decreases with increasing temperature until at 14000° K. the content has fallen to 0.27%.

The results of these computations are shown in table 5. The temperature T_2 in this case is the independent variable.

At temperatures below 4000°K . the contribution made by the energy of dissociation to the internal energy is small, but at 14000°K . it is as much as 60%. It is computed by assessing the energy liberated by the initial dissociation of the nitrogen and oxygen molecules and subtracting a certain amount due to the recombination of some of the atoms to form nitric oxide.

Table 5. Shock wave conditions in air

$T_2 (^{\circ}\text{K.})$	4,000	6,000	8,000	10,000	14,000
x	0.1571	0.7298	0.9257	0.9678	0.9917
y	0.000092	0.0092	0.0885	0.2968	0.7809
z	0.1025	0.0875	0.0430	0.0226	0.0061
% mass in air of					
NO	4.46	3.81	1.87	0.98	0.27
O	3.64	16.93	21.48	22.45	23.01
O ₂	17.18	4.24	0.73	0.22	0.05
N	0.01	0.70	6.68	22.41	58.96
N ₂	73.41	73.02	67.94	52.64	16.41
A	1.30	1.30	1.30	1.30	1.30
ΔE (cal/gm.)	1040	2024	3090	4790	8540
p_2 (atmos.)	127	236	366	561	1003
v_2 (cm ³ /gm.)	91.7	81.8	78.4	72.5	71.9
U (m/sec.)	3350	4540	5645	6965	9310
u (m/sec.)	2950	4060	5075	6320	8445

In table 6 the shock wave variables are given with the pressure as the independent variable.

Table 6

p_2 (atmos.)	200	400	600	800	1,000
U (m/sec.)	4,180	5,890	7,200	8,310	9,310
u (m/sec.)	3,750	5,300	6,550	7,560	8,450
$T_2 (^{\circ}\text{K.})$	5,270	8,420	10,300	12,000	14,000
E (cal/gm.)	1,730	3,280	5,100	6,670	8,420
v_2 (cm ³ /gm.)	82.7	76.8	71.8	69.9	71.8

The U and u values are not far removed from those given in table 4 as the first solution. The T_2 figures, however, have changed a good deal, and at $p_2=500$ atmospheres T_2 has increased by 1480°K . The ΔE values are still in close proximity and show a practically linear dependence on p_2 . There is now a minimum value of v_2 (about 70 cm^3) at $p_2=800$ atmospheres instead of at 600 atmospheres.

§ 7. NORMAL IMPACT OF A SHOCK WAVE ON A RIGID WALL

The pressure experienced by a rigid wall when a shock wave strikes it is many times greater than the shock wave pressure p_2 . The simple theory is quite adequate for low values of pressure, but for the higher values calculation becomes difficult. The method makes use of the fact that the air behind the returning shock wave must be at rest. Thus, if p_2 and v_2 are the pressure and volume of unit mass of air before impact, p_3 and v_3 similar quantities after impact, u the mass-velocity produced immediately in the rear of the shock wave by p_2 and v_2 , then

$$v_3 = v_2 - u^2/(p_3 - p_2). \quad \dots\dots(29)$$

Outside the "dissociation region" solution is easy. Assume a value for p_3 , then equation (29) gives v_3 and hence T_3 by the perfect-gas laws. The internal energy of air function, $E(p, T)$, has to be plotted as a function of temperature for equal values of p_3 —considerably greater in magnitude than was required for p_2 . Interpolation yields $(E_3 - E_2)$ where E_2 and E_3 are the energies of air at the shock wave before and after impact. If these values agree with that obtained from the formula

$$E_3 - E_2 = \frac{1}{2}(p_3 + p_2)(v_2 - v_3), \quad \dots\dots (30)$$

then the p_3 chosen is the correct one. Normally p_3 has to be readjusted so that this requirement is satisfied.

Inside the "dissociation zone" solution is more difficult, as $v_3(p, T)$ has to be plotted for equal pressures (figure 8). The steps are :

- (i) Assume a certain p_3 value, equation (29) gives v_3 ;
- (ii) plot 8 (b) gives the associated T_3 ;
- (iii) plot 8 (a) gives E_3 ;
- (iv) hence $E_3 - E_2$ (known) $= \Delta E$;
- (v) compare with value given by equation (30);
- (vi) adjust p_3 to obtain agreement.

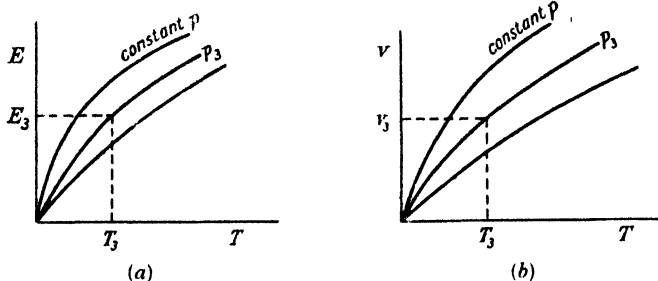


Figure 8.

In the absence of NO and argon, and with the old value of D for nitrogen, the following values for p_3/p_2 have been evaluated.

Table 7

p (atmos.)	15	30	50	65	100	200	300	375	500	800	1000
p_3/p_2	5.87	6.83	7.8	8.2	8.9	10.3	12.8	13.8	14.0	12.3	10.8

It is unlikely that correction for formation of NO, etc., will affect these values to any appreciable extent.

§ 8. EFFECT OF IONIZATION

The method used for investigating the possible changes in the internal energy, equilibrium constants and shock wave properties, due to ionization in the earth's surface atmosphere due to the very high temperatures considered in this paper, is similar to that for evaluating the "electronic" contributions. Possible ionic formations are accounted for by a similar addition to the partition functions and yield contributions as follows:

Ionic contribution to the internal energy

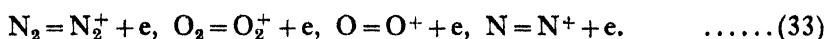
$$E_i = \{i_1 I_1 \exp(-I_1/kT)\} / \{g_0 + i_1 \exp(-I_1/kT)\}; \quad \dots\dots (31)$$

contribution to the previous $\ln K$ formula.

$$\ln \left\{ \frac{\left[1 + \frac{i_1^A}{g_0^A} \left(\exp - \frac{I_1^A}{kT}\right)\right] \left[1 + \frac{i_1^B}{g_0^B} \left(\exp - \frac{I_1^B}{kT}\right)\right]}{\left[1 + \frac{i_1^{AB}}{g_0^{AB}} \left(\exp - \frac{I_1^{AB}}{kT}\right)\right]} \right\}, \quad \dots\dots (32)$$

where i_1 is the statistical weight of atoms (or molecules) in the main ionized state and I_1 the energy difference from ground state to ionized state.

These equations cover the transformations



The spectroscopic data required are taken from *Molecular Spectra and Molecular Structure of Diatomic Molecules* by Herzberg, and *Atomic Energy States* by Bocher and Goudsmidt.

	Electronic potentials (ev.)	State	Statistical weights
O^+	13.60	^4S	4
N^+	14.54	^3P	1
O_2^+	12.20	$^2\Pi_g$	2
N_2^+	15.58	$^2\Sigma^+$	2

These electronic potentials are very considerably higher than the excited levels of the un-ionized gases. Indeed, substitution of these numerical values in equations (31) and (32) shows immediately that ionic formation at high temperatures has a negligibly small effect on the internal energies and K values. It does not, therefore, affect the shock wave properties.

For temperatures greater than 16000°K . ionization plays an increasing part; the calculations to allow for this are described by Dr. Kynch in his papers in the "Atomic Energy" series. The problem at these very high temperatures is eased by complete dissociation.

ACKNOWLEDGMENTS

The author wishes to thank the Chief Scientific Adviser, Ministry of Home Security, for permission to publish this paper, and Dr. W. G. Penney for many helpful discussions.

REFERENCE

LEWIS, B., and VON ELBE, G., 1938, *Combustion, Flames and Explosion of Gases* (Cambridge: University Press), pp. 239-244.

DISCUSSION.

Dr. A. H. DAVIS. The present paper deals with the relations which exist between pressure, wave velocity, particle velocity, etc., whereby the observations of one quantity at a point—say of the velocity of the wave—makes it possible to infer the magnitude of others—say of the pressure, temperature cycle and so on. It does not deal with the propagation of shock waves through the atmosphere. Moreover, whilst theoretical work has been conducted on the propagation of sounds of "small" amplitude through the atmosphere, around obstacles and so on, I believe that there is not yet any theory of the propagation of finite shock waves under similar conditions, and consequently an interesting theoretical field lies open.

The genesis of a shock wave may be observed by exploding an explosive mixture of hydrogen and oxygen contained in an india-rubber balloon of about 6 cubic feet in volume. (The operations of filling the balloon, first partially with one gas, and then finally with the other, must be conducted by remote control methods, as the balloon is liable to detonate when handled under some atmospheric conditions.) Observations of the blast pressure at distances of some 3 ft. to 10 ft. from the explosion reveal a wave which is first rather non-descript, but which develops into a shock wave after travelling a few feet.

Mr. A. WARREN. I should like to draw attention to some unusual features in the shape of the shock waves recorded near the ground with small high explosive charges.

The measurements were made (using an experimental technique recently described by Grime and Sheard (*Proc. Roy. Soc. A*, 1946, **187**, 357) up to a distance at which the calculated

velocity of the shock wave is within $\frac{1}{4}\%$ of the normal velocity of sound, and under these circumstances it was anticipated that any change in the wave form of the shock pulse would be insignificant.

It was found that up to a certain distance a wave-form characteristic of normal shock propagation was recorded, but that subsequently a change of waveform occurred; a secondary pressure peak commenced to form behind the shock front and the two peaks tended to separate. Since the maximum excess pressure in the initial shock front decayed at an abnormally high rate, the maximum excess pressure in the secondary peak soon exceeded that in the shock front, and in some instances the initial shock front almost disappeared at the greater distances.

It was observed that the critical distance beyond which the secondary peak pressure exceeded that in the shock front appeared to be greater when the propagation losses in the wave front were reduced. For example, it was greater (a) in the direction of the prevailing wind, (b) for gauges at (say) 3 feet above ground as compared with similar measurements very near the ground, and (c) if the surface of the ground between the charge and gauge was covered with sheets of smooth material, e.g. strawboard with improved reflection characteristics.

It was known from the recent series of papers by Friedlander ("Diffraction of sound pulse", *Proc. Roy. Soc. A*, 1946, **186**, 322) that after diffraction from obstacles of various types a shock wave was likely to lose sharpness and become broadened. A similar effect was anticipated in an unpublished mathematical treatment by Prof. Sir. G. I. Taylor dealing with the effects due to turbulence, and to the gradient of wind velocity near the surface of the ground.

It appeared from the records already obtained that modified wave-forms very similar to those to be anticipated on theoretical grounds from the effects of diffraction or turbulence were observed, but that in addition, under certain circumstances, this modified wave-form could co-exist with an initial shock front of characteristic shape at an appropriate time separation.

MR. A. R. BRYANT. I should like to ask what the temperature would be if an explosive were surrounded by an envelope of argon instead of air. One would expect this to be much higher owing to argon having a higher value of γ and owing to absence of dissociation. With regard to the results of the shock wave calculations could the author state how sensitive the quantities behind the shock front are to the assumption that thermodynamic equilibrium is or is not attained?

AUTHOR'S reply. I have read with interest the remarks of Dr. A. H. Davis and Mr. A. Warren. An approximate solution of the problems of hydrodynamical flow behind a shock wave set up in air as a result of an explosion from a spherical charge was, in fact, attempted in 1941 by Dr. W. G. Penney and myself (unpublished). At that time the adiabatics of air at high temperature and pressure had not been evaluated, and γ was taken to be equal to 1.403. This assumption was likely to introduce large numerical errors, but the results showed the presence of a secondary pressure peak, as mentioned by Mr. Warren. The problem was investigated during 1943-45 in the U.S.A., but I am not aware if any of this work has yet been published.

In reply to Mr. Bryant's first question, I agree that the temperature would be higher if the explosive were surrounded by argon of the same density as air in the normal atmosphere, the internal energy of argon per gramme at a given temperature being less than that of air; there is no rotational, vibrational energy, and no dissociation to be allowed for.

With regard to Mr. Bryant's second question, the shock wave velocity and pressure are probably very little affected by the assumption of thermodynamic equilibrium. The actual temperature, however, will be higher than those obtained on this assumption. In the paper it is assumed that the energy is taken up in the translational, rotational and vibrational forms, distributed according to the equations of statistical mechanics. A given volume of air receives a definite quantity of energy from the "shock wave", and it is probable that, due to the nature of the "shock", rather more energy is taken up into the translational type and less into the rotational and vibrational types than given by the theory. Hence the temperature would be higher than expected on the assumption of thermodynamic equilibrium.

The Reflection of a Plane Shock Wave at a Gaseous Interface

By STEWART PATERSON

Imperial Chemical Industries Ltd, Explosives Division, Stevenston, Scotland

MS. received 18 February 1948

ABSTRACT. A plane shock wave falls normally on the interface between two ideal gases of molecular weights M , m and constant ratios Γ , γ of specific heats. As a rule, the character of the reflected wave then depends only on the ratio $\Gamma M/\gamma m$. In exceptional cases, however, the type of reflection may change at a critical value of the incident intensity. An example of this is given.

Two perfect gases with specific volumes V_0 and v_0 and constant ratios Γ , γ of specific heats are at rest and separated by an ideally thin and light partition. The pressure is p_0 and temperature T_0 in each gas. A plane shock wave S_1 , the steady conditions behind which are indicated by suffix 1, passes through the first gas and falls normally on the interface.

A wave S_2 will then be transmitted through the second gas; we denote conditions behind S_2 by a suffix 2. These conditions are defined by the familiar Rankine-Hugoniot relation if S_2 is a shock wave, or by the Riemann relation for a progressive wave if S_2 is a rarefaction. Thus, if W represents in general the mass velocity behind a wave (measured in the direction of propagation of S_1), and π the pressure ratio across it, so that $\pi_1 \equiv p_1/p_0$, $\pi_2 \equiv p_2/p_0$, we have for S_1

$$W_1 = \sqrt{\{(\Lambda - 1)p_0 V_0\}(\pi_1 - 1)/\sqrt{(\Lambda\pi_1 + 1)}}, \quad \dots\dots(1)$$

and for S_2

$$W_2 = \begin{cases} \sqrt{\{(\lambda - 1)p_0 v_0\}(\pi_2 - 1)/\sqrt{(\lambda\pi_2 + 1)}}, & \text{if } \pi_2 > 1 \\ -\sqrt{\{(\lambda^2 - 1)p_0 v_0\}[1 - \pi_2^{1/(\lambda+1)}]}, & \text{if } \pi_2 < 1 \end{cases} \quad \dots\dots(2)$$

where $\Lambda \equiv (\Gamma + 1)/(\Gamma - 1)$ and $\lambda \equiv (\gamma + 1)/(\gamma - 1)$. It is of course physically obvious, and will be shown below, that S_2 must be a shock wave ($\pi_2 > 1$) and not a rarefaction wave ($\pi_2 < 1$). However, for the moment we need not assume this.

A reflected wave S_3 will also arise in the first gas; we denote conditions behind by a suffix 3, so that $\pi_3 \equiv p_3/p_1$. S_3 may be either a shock wave ($\pi_3 > 1$) or a rarefaction wave ($\pi_3 < 1$), and the mass velocity behind it is

$$W_3 = \begin{cases} W_1 - \sqrt{\{(\Lambda - 1)p_1 V_1\}(\pi_3 - 1)/\sqrt{(\Lambda\pi_3 + 1)}}, & \text{if } \pi_3 > 1 \\ W_1 + \sqrt{\{(\Lambda^2 - 1)p_1 V_1\}[1 - \pi_3^{1/(\Lambda+1)}]}, & \text{if } \pi_3 < 1. \end{cases} \quad \dots\dots(3)$$

Continuity of velocity and pressure across the interface requires, however, that $W_2 = W_3$ and also $p_2 = p_3$, so that $\pi_2 = \pi_3\pi_1$. Making these substitutions in (2) and using (3) and (1), we are in a position to determine π_3 , and so all the properties of the reflected and transmitted waves. The procedure is now well known.

It may easily be shown that W_2 , as given by (2) with $\pi_2 = \pi_3\pi_1$, is a monotonic function of π_3 which increases continuously through zero as π_3 increases through $1/\pi_1$. In the same way, W_3 is a monotonic function of π_3 which decreases continuously through W_1 as π_3 increases through unity. It therefore appears at once, since $W_1 > 0$,

- (a) that π_3 cannot be less than $1/\pi_1$; in other words, that π_2 must always exceed 1, so that the transmitted wave is necessarily a shock wave;

(b) that π_3 may be either greater than or less than 1, so that the reflected wave may be of either type, according to circumstances.

These properties are illustrated in figures 1 and 2 by the case of a 2-atm. shock wave proceeding from oxygen to argon and vice versa, both gases being initially at 1 atm. and 273° K.

The condition that S_3 be a shock wave is that the two curves intersect at a point for which $\pi_3 > 1$; but in view of the above remarks concerning the course of $W_2(\pi_3)$ and $W_3(\pi_3)$, this condition can be replaced by the analytically much simpler one that $W_3(\pi_3) > W_2(\pi_3)$ when $\pi_3 = 1$; in other words, that

$$(\Lambda - 1)V_0/(\Lambda\pi_1 + 1) > (\lambda - 1)v_0/(\lambda\pi_1 + 1), \quad \dots\dots(4)$$

which may be written

$$(\pi_1 - 1)[(\Gamma + 1)\Delta_0 - (\gamma + 1)\delta_0] < 2[\gamma\delta_0 - \Gamma\Delta_0], \quad \dots\dots(5)$$

where $\Delta_0 V_0 \equiv 1 \equiv \delta_0 v_0$.

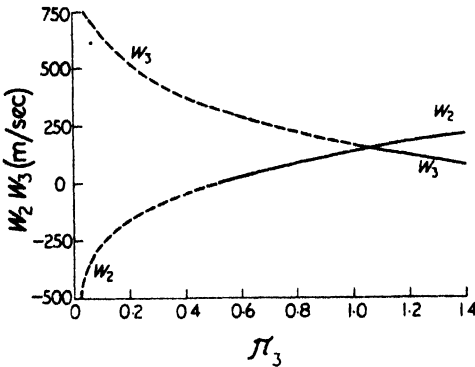


Figure 1. Reflection of a 2-atm. shock wave in oxygen by argon. $p_0 = 1$ atm.; $T_0 = 273^\circ$ K.; W_2, W_3 = mass velocities behind transmitted and reflected waves; π_3 = pressure ratio across reflected wave, which is seen to be a shock.

———— Shock
----- Rarefaction

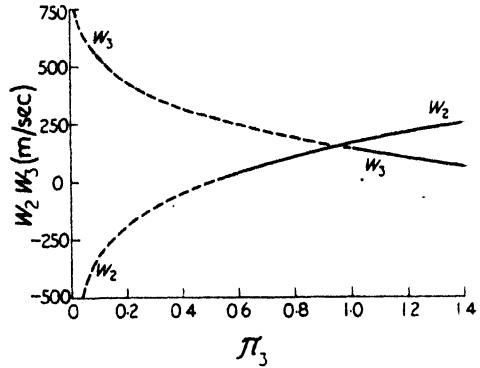


Figure 2. Reflection of a 2-atm. shock wave in argon by oxygen. $p_0 = 1$ atm.; $T_0 = 273^\circ$ K.; W_2, W_3 = mass velocities behind transmitted and reflected waves; π_3 = pressure ratio across reflected wave, which is seen to be a rarefaction.

———— Shock
----- Rarefaction

Four alternatives now arise:

(a) $\Gamma\Delta_0 < \gamma\delta_0$ and $(\Gamma + 1)\Delta_0 < (\gamma + 1)\delta_0$.

Then (5) is satisfied irrespective of $\pi_1 (> 1)$, and the reflected wave is always a shock wave.

(b) $\Gamma\Delta_0 > \gamma\delta_0$ and $(\Gamma + 1)\Delta_0 > (\gamma + 1)\delta_0$.

Then (5) cannot be satisfied for any $\pi_1 (> 1)$, and the reflected wave is always a rarefaction wave.

(c) $\Gamma\Delta_0 < \gamma\delta_0$ but $(\Gamma + 1)\Delta_0 > (\gamma + 1)\delta_0$.

Then (5) is satisfied for $\pi_1 < \pi_1'$, but not for $\pi_1 > \pi_1'$, where

$$\pi_1' \equiv \frac{(\gamma - 1)\delta_0 - (\Gamma - 1)\Delta_0}{(\Gamma + 1)\Delta_0 - (\gamma + 1)\delta_0}. \quad \dots\dots(6)$$

Thus, for $\pi_1 < \pi_1'$, S_3 is a shock wave, but for $\pi_1 > \pi_1'$, S_3 is a rarefaction wave.

(d) $\Gamma\Delta_0 > \gamma\delta_0$ but $(\Gamma + 1)\Delta_0 < (\gamma + 1)\delta_0$.

Then, for $\pi_1 < \pi_1'$, S_3 is a rarefaction wave, but for $\pi_1 > \pi_1'$, S_3 is a shock wave.

Cases (a) and (b) are evidently reciprocal. Thus they show that if shock waves in the first gas are always reflected as shock waves by the second, shock waves in the second will always be reflected as rarefaction waves by the first, and vice versa. This is physically reasonable. Cases (c) and (d) are reciprocal in a similar sense, but the existence of a critical incident pressure ratio π_1' , at which the reflection changes type, is more surprising. It remains to be seen whether this effect can be realized in real gases. The necessary condition may be written, for convenience,

$$\Gamma M \leq \gamma m, \quad (\Gamma + 1)M \geq (\gamma + 1)m, \quad \dots\dots(7)$$

where M , m denote molecular weights. Examination of tabulated specific heats and molecular weights shows at once that (7) is not usually satisfied, so that cases (a) and (b) form the rule. However, the possibility of satisfying (7) is not excluded if Γ and γ are sufficiently different. Suppose, for example, that $\Gamma = 1.1$ and $\gamma = 1.667$. Then (7) requires that $1.27m < M < 1.515m$. It appears in fact that (7) is satisfied by the following pairs of gases: neon with acetylene, argon with butane vapour or methyl chloride vapour, for which approximate constants at 1 atm. and 273°K. and corresponding values of π_1' are given in the following table:

Gas	Γ	Δ	$\Gamma\Delta$	$(\Gamma+1)\Delta$	π_1'
Neon (Ne)	1.642	0.966 (g/l.)	1.586	2.553	(with Ne)
Acetylene	1.28	1.190	1.52	2.71	1.78
Argon (A)	1.667	1.781	2.964	4.750	(with A)
Butane vapour	1.11	2.599	2.88	5.48	1.51
Methyl chloride vapour	1.279	2.238	2.862	5.10	1.60

Calculations for the vapours are rather unreliable, in view of the high critical temperatures and correspondingly large departures from the ideal gas laws. For neon and acetylene, however, the effect is probably authentic, and could perhaps be detected in Schlieren photographs by the reversal of the sign of $\pi_3 - 1$ at π_1' ; the wave- and mass-velocities do not, of course, change sign. The properties of transmitted and reflected waves in this system up to $\pi_1 = 2.5$ are shown in figure 3.

For the great majority of gaseous systems (7) fails, and the condition for a reflected shock is then simply $\Gamma M < \gamma m$, which may also be written $A_0 M < a_0 m$ or

$$\left| \left(\frac{\partial V_0}{\partial p_0} \right)_s \right| < \left| \left(\frac{\partial v_0}{\partial p_0} \right)_s \right|,$$

where A_0 , a_0 denote velocities of sound, and S entropy.

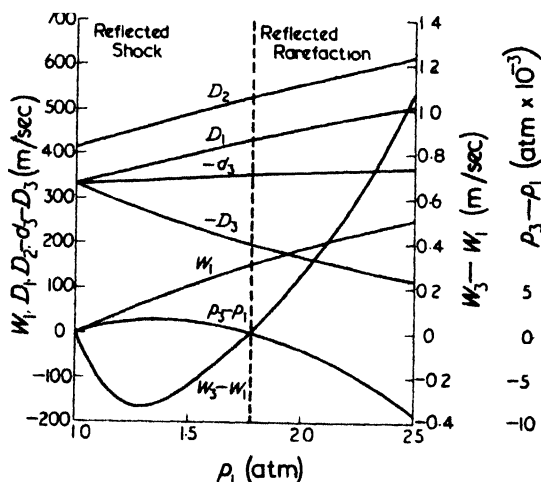


Figure 3. Reflection of weak shock waves in acetylene by neon. $p_0 = 1$ atm.; $T_0 = 273^\circ$ K.; p_1 , p_3 = pressures, and W_1 , W_3 mass velocities, behind incident wave and between transmitted and reflected waves respectively; D_1 , D_3 , D_3 = absolute velocities of incident, transmitted and reflected waves; d_3 = velocity of reflected wave relative to gas behind incident wave. Velocities are positive when in the direction of incident wave.

A Study of Transient Radar Echoes from the Ionosphere

By E. EASTWOOD AND K. A. MERCER

Communicated by E. V. Appleton ; MS. received 21 February 1947 and in amended form 22 November 1947; read 31 January 1947

ABSTRACT. The rate of occurrence of transient bursts of ionization within the lower E region of the ionosphere has been systematically measured during the period January 1945 to July 1946.

It has been found that the activity varies both diurnally and seasonally and in such a manner as to lend support to the meteor theory of burst formation. Some measure of solar control of the rate of burst occurrence has also been detected ; this effect may be explained in terms of the normal diurnal E region changes in density of ionization. From experiments performed during the solar eclipse of July 1945, it is concluded that the sun does not emit a burst-producing radiation.

Analysis of the observations suggests that the greater proportion of the bursts are created within a thin layer located at a height of 86 km. ; the distribution of the bursts within this layer has proved to be uniform over wide areas and no latitude effect has been detected. It is established that the rate of incidence of bursts which present echoing areas between A and $A + dA$ m² to a radio wave of frequency ν may be expressed in the form $CdA/\nu^3 A^{3/2}$, with C constant.

§ 1. INTRODUCTION

THE density of ionization within the E region of the ionosphere normally varies in a regular manner with the intensity of the ultra-violet light received from the sun, but a number of workers have shown the temporary existence within the layer of regions of much higher density. These disturbances appear to be of two distinct types. Appleton and Naismith (1935, 1940) describe one form of this sporadic ionization that has come to be known as "abnormal or intense E"; this type usually persists for minutes or even hours and may extend over large areas. A second type of E-layer disturbance was also recognized by Appleton, Naismith and Ingram (1937) in the form of localized bursts of ionization whose density greatly exceeded that characteristic of abnormal E. These ion clouds were observed to occur both by day and by night and yielded transient echoes of only about a second duration. The reflection coefficients and heights of occurrence of these clouds were later studied in more detail by Appleton and Piddington (1938). It is with this second type of E-layer disturbance, or "ionospheric burst" as it is now commonly termed, that the present paper is concerned.

The problem of E-layer disturbances was examined by Eckersley (1940) from the point of view of radio communication effects and he has shown that the presence of signals within the skip zone is attributable to scattering from ionospheric "clouds". The rate of occurrence of the clouds was found to follow a diurnal cycle, while the dependence of the intensity of the scattered radiation upon the wavelength was in agreement with a fourth power law previously suggested by him (1932).

The nature of the ionizing agent responsible for these E-layer perturbations is a question which has given rise to speculation, but much of the evidence now

points towards the validity of the meteor theory. Nagaoka (1929) was the first to consider the possible influence of meteors on radio transmission and suggested that disturbed ionospheric conditions might arise from the removal of electrons by a process of attachment to particles of meteoric dust. Skellett (1932, 1935) disagreed with this theory and, supported by experimental work performed in collaboration with Schafer and Goodall (1932), showed that impact ionization by meteors was a more likely explanation of the observed phenomena. Pierce (1938) also found that sudden increases in the abnormal E type of ionization could be explained by the passage overhead of a single large meteor, while certain experiments reported by Eckersley (1937) on the occurrence of scattering clouds were considered by Skellett (1938) to be consistent with the meteor theory of their formation. Further convincing evidence in support of the meteor hypothesis has recently been provided by Appleton and Naismith (1946), and by Hey and Stewart (1946), but that certain difficulties still remain in applying the theory is shown by the work of Eckersley and Farmer (1945) on the rapid variation of the direction of arrival and state of polarization of the waves scattered from the bursts.

The aim of the present investigation has been to gather further information upon the characteristics of ionospheric bursts by making a systematic study of the time variation of burst activity. Observations were also made during the total solar eclipse of July 1945 to detect any influence which the sun might exert upon the phenomenon of burst occurrence.

§ 2. EXPERIMENTAL METHODS

The "pulse-echo" technique has been used in the present work for the detection and location of ionospheric bursts. The experiments have been conducted at a number of Royal Air Force radar stations each of which was equipped with a high power, pulsed radio transmitter and a sensitive receiver. The transmitter aerial consisted of a vertical stack of horizontal, half-wave dipoles and was situated within 200 metres of a receiver aerial with dipole elements parallel to and at the same mean height as those of the transmitter. The radio-frequencies employed ranged from 20 to 45 Mc/sec. but the pulse recurrence frequency was 25 pulses/sec. at all the stations.

The echo signal returned by the burst was recorded photographically in the usual ($P' t$) type of record. With the aerial system described above most of the energy is radiated at comparatively low angles of elevation; a method of oblique incidence has therefore been used, and the P' of the records represents not the usual equivalent height of the ionospheric reflecting layer but the slant range of the ionospheric burst. The magnitudes of the echoes were also recorded photographically using a ciné camera which was operated without its shutter. It was found that successive sweeps on the cathode-ray display tube were registered as separate parallel traces on the film so that the echo amplitudes could easily be measured from the trace deflections. This simple technique permitted the rates of growth and decay of particular responses to be studied.

§ 3. DIURNAL AND SEASONAL VARIATIONS

In order to investigate diurnal and seasonal changes in the rate of burst occurrence observations were made from two identical stations at Bawdsey, Suffolk, on a frequency of 22.69 Mc/sec., a pulse width of 8 μ sec. and a power

of 1 Mw. The hourly rate of occurrence of bursts formed within a selected slant range interval was measured over periods of twenty-four hours. The observations covered the period January 1945–July 1946, with the 24-hour sequence of observations taking place twice each week.

Every precaution was taken to ensure constant performance of the equipment but output power variations did occur and, in order to apply appropriate correction, the relationship between the number of bursts observed and the power radiated was separately investigated. This work established the result that, for a given system and range interval, the number of bursts observed was directly proportional to the square root of the power. From this result it was now possible to reduce the average measured power over an interval to the standard power of 1 Mw. and to correct the burst rate accordingly.

The number of bursts recorded during the successive hours of a day's run clearly showed the presence of a diurnal cycle of burst activity. These daily curves possessed certain features in common but marked changes also occurred from day to day and curves of mean values for each month were therefore plotted (figures 1 and 2). The curves of figure 1 show relative monthly changes in the daily cycle; the absolute levels of activity are not comparable since the same aerial system was not used from month to month. The curves of figure 2, for which the same aerial system was used, are however, strictly comparable both for total activity and for relative monthly changes.

Figures 1 and 2 show clearly the presence of a diurnal cycle of burst activity and also show that the form of the daily cycle changes with the time of the year. The diurnal cycle is characterized by three main features—a pronounced minimum in the late afternoon, and maxima in the early and late morning. It will be seen that the minimum of the daily curve is the most clearly defined and persistent feature of all the histograms and that its time of occurrence is approximately constant and independent of the time of sunset. The two maxima of the daily curve occur at

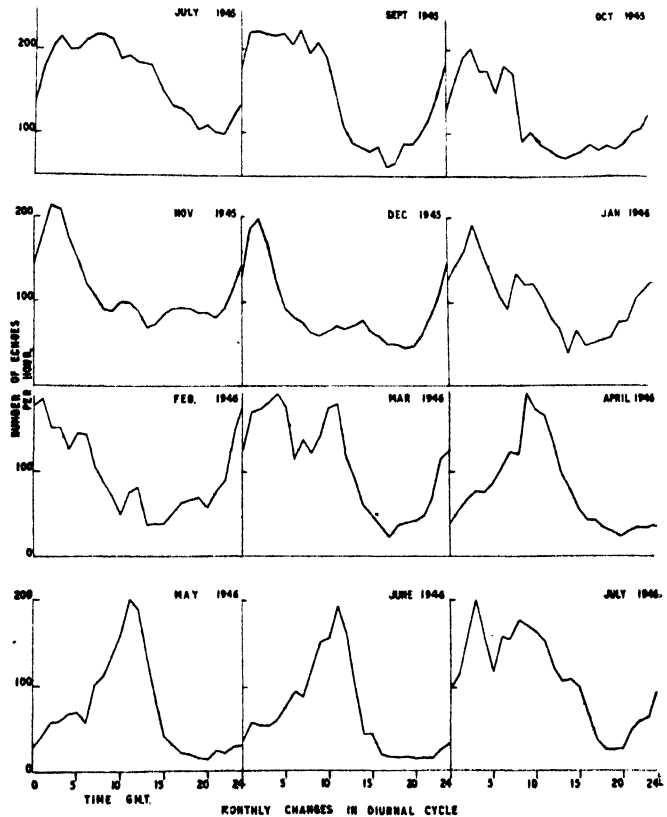


Figure 1.

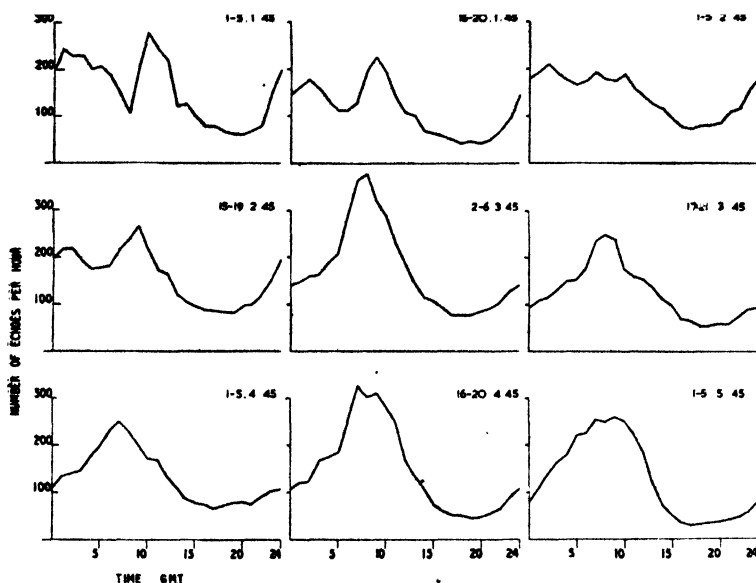


Figure 2.

approximately 0400 and 1200 U.T., and do not appear to vary with time of sunrise and sunset through the year. The relative amplitudes of the two maxima are, however, dependent upon the season of the year; in particular, the early morning peak tends to be associated with the period of the winter solstice while the late morning peak becomes the dominant feature of the curve about the time of the summer solstice. This is seen in the middle column of curves in figure 1: December 1945 and June 1946 show pronounced early and late morning peaks respectively, while September 1945 and March 1946 have transitional types of curves with both peaks present and of roughly the same amplitudes.

The change in total volume of daily burst activity throughout the year was studied by taking observations on convenient days when the same aerial system could be used. The total number of echoes during a 24-hour period observed within the range interval 160-320 km. was determined and the average number of responses per hour extracted. Figure 3 shows these average daily figures

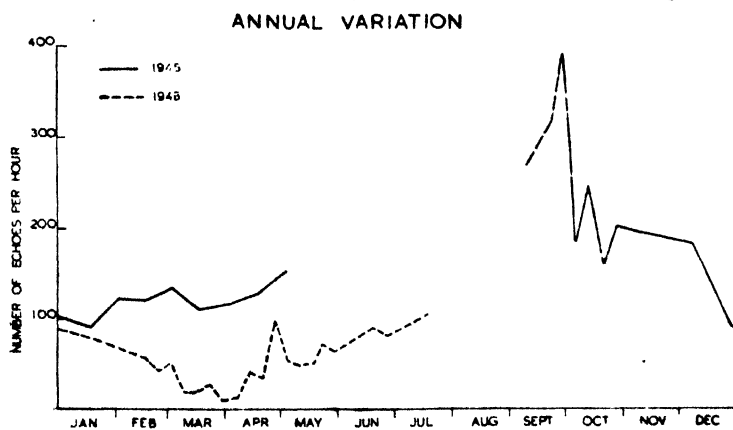


Figure 3.

plotted against the appropriate dates of observation. It is seen that a seasonal cycle of activity is present with its maximum and minimum occurring about the times of the autumn and spring equinoxes respectively.

§ 4. SOLAR ECLIPSE EXPERIMENTS

An experimental investigation of solar effects on ionospheric burst formation was possible during the total solar eclipse of 9th July 1945. Such solar effects might be expected to arise in two ways, either directly due to the emission of a burst-producing radiation from the sun or indirectly from the sun's control of the ionized state of the ionosphere. In both these cases the rate of occurrence of the bursts would be the feature most likely to change during the course of a total eclipse of the sun. The general plan of the experiments, therefore, was to compare the rate of formation of the bursts on the eclipse day with the average rate obtained from observations made during two control periods each of seven days, on either side of the eclipse day.

The experiments were conducted from two groups of stations in the north of Scotland, lying within the belt of totality for a particle radiation at the range of velocities to be expected. The activity throughout the eclipse period was measured in terms of the number of echoes observed per half hour and was compared with the mean activity derived from a grouping of the corresponding periods of the control days. The amplitudes and durations of echoes observed during the eclipse were similarly compared with the averages derived from the control periods. This procedure was followed at all the stations of table 1.

A careful statistical analysis of all the results obtained failed to reveal any significant departure of ionospheric burst characteristics observed during the period of the eclipse from the averages established by the control periods. It may be concluded therefore, that the sun is not a source of burst-producing particles. However, the great volume of data obtained in the course of these experiments permitted some interesting conclusions to be drawn on the questions of time variation and space distribution of burst activity. Figure 4 presents a convenient summary of the daily burst activity recorded at the various stations during the sixteen days of the experiments; it will be noticed that the activity at any particular station changes greatly from day to day and that these

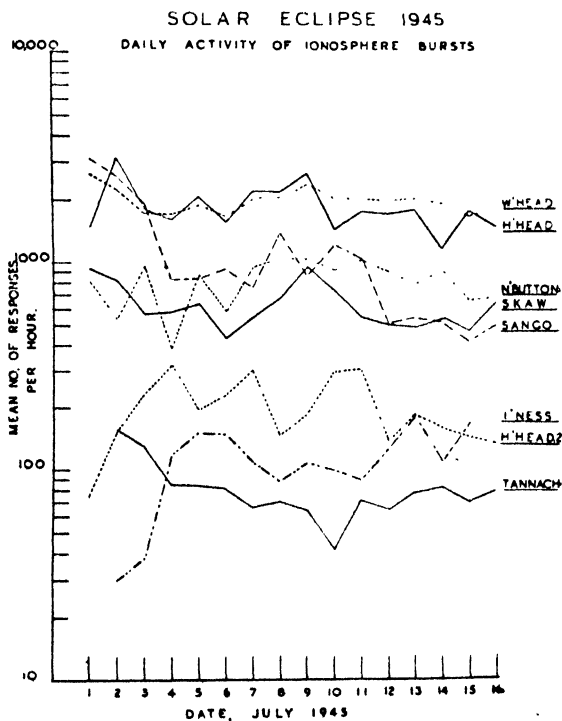


Figure 4.

changes occur in an apparently random manner at the different stations. The large differences between the mean activity levels of the stations taken over the whole duration of the observations are particularly noteworthy and will be discussed in more detail later.

§ 5. RANGE DISTRIBUTION OF BURSTS

From the ($P't$) records obtained during the eclipse experiments a histogram was derived for each of the stations relating the number of echoes in successive 16 km. intervals of range with the mean ranges of these intervals. It was found that every station possessed a characteristic range-distribution histogram. Similarly, each of the various aerial systems used at Bawdsey was found to have a characteristic histogram; figure 5 shows a typical histogram secured from a two-stack receiver aerial system combined with an eight-stack transmitter array, both aeriels centred on a mean height above sea-level of 255 ft. and operated on a frequency of 22.69 Mc/sec. Figure 6 shows the distribution corresponding to a single dipole receiver aerial and a single dipole transmitter, each at a height of one-quarter wavelength above the ground.

In all these experiments a sharp cut-off in range was observed at about 80 km., in agreement with the accepted view that the bursts are formed within the E

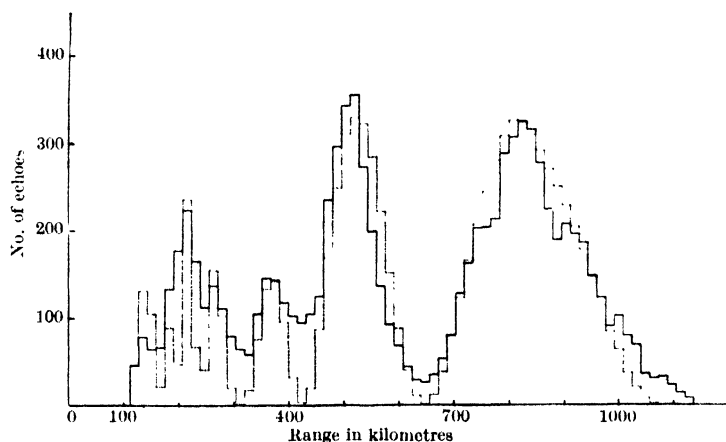


Figure 5. Range distribution.

———— Bawdsey. Theoretical.

layer, but it was also noticed that echoes occasionally appeared at less than this range. Detailed analysis has shown that the form of a station's range-distribution histogram may be convincingly explained in terms of the interaction of the station's characteristic radiation pattern with a burst-producing layer situated within the lower E region of the ionosphere. The sharpness of certain of the histogram minima, e.g. that at 650 km. in figure 5, suggests that the bursts are probably confined to a very thin layer so that, to a first approximation, we may neglect a height distribution of bursts and seek for a theoretical development of the histograms on the assumption that the bursts are distributed uniformly over a spherical surface.

The number of bursts occurring within the range interval R to $R+dR$ and having equivalent echoing areas between A and $A+dA$ square metres will be proportional to the zone of the ionosphere intercepted between these ranges and also to the occurrence rate per unit area of bursts of this magnitude, say

ndA . If h is the height of the burst-producing layer and ρ be the radius of the earth, then the area of the zone is $2\pi(1 + h/\rho)R dR$, and we may therefore write ;

Number of bursts observed in the range
interval R to $R + dR$ having echoing areas $\propto n R dA dR$, ...(1)
between A and $A + dA$:

i.e., total number of bursts observed in
the range interval R_1 to R_2 : $\propto \int_{R_1}^{R_2} \int_{A_L}^{\infty} n R dA dR$(2)

In this expression A_L is the smallest echoing area which can be detected by the given equipment at the range R and it will be shown later (equation (9)) that

$$\int_{A_L}^{\infty} n dA \propto 1/\sqrt{A_L},$$

so that (2) may be rewritten :

Total number of bursts observed in the
interval R_1 to R_2 : $\propto \int_{R_1}^{R_2} R A_L^{-1/2} dR$(3)

Now a burst of echoing area $A m^2$ formed at a range R from a station radiating a power of P kilowatts on a wavelength of λ m. will return an echo whose signal/noise ratio is Z , where Z is given by

$$Z \propto \lambda R^{-2} \sqrt{(A P n_t n_r)} \phi(\alpha), \quad \text{.....(4)}$$

in which n_t and n_r are the numbers of dipoles composing the transmitter and receiver aerials respectively and $\phi(\alpha)$ is the product of the interference and grating factors of the arrays, expressed as a function of the angle of elevation α . For a given elevation this function varies with the azimuth, due to the form of the radiation pattern from a half-wave dipole, and it is therefore necessary to work with the azimuthal average value of $\phi(\alpha)$.

Since the smallest echo signal which can be distinguished from the noise corresponds to a constant value of Z , it follows from (4) that the least echoing area, A_L , which can be detected at a range R and elevation α is expressed by

$$A_L^{-1/2} \propto \lambda R^{-2} \sqrt{(P n_t n_r)} \phi(\alpha). \quad \text{.....(5)}$$

Substituting this value of A_L in (3) gives

Total number of bursts observed in the range interval R_1 to R_2

$$\propto \int_{R_1}^{R_2} \lambda R^{-1} \sqrt{(P n_t n_r)} \phi(\alpha) dR, \quad \text{.....(6)}$$

which may be expressed

$$\propto \int_{R_1}^{R_2} Q(R, h) dR, \quad \text{.....(7)}$$

where h is the height of the layer of burst formation.

In order to develop the range distribution histogram for a particular system the appropriate $Q(R, h)$ was evaluated and plotted against R for a selected value of h , suitable correction being applied for the effects of atmospheric refraction. The resulting curve was integrated over successive 16 km. range intervals by use of Simpson's Rule and the areas of the consecutive blocks compared with the experimental histogram as in figure 5, where the chain lines show the theoretical histogram derived for the Bawdsey system. Theoretical histograms were similarly calculated for all the stations and for a number of selected values of h ; comparison of the theoretical and experimental patterns showed that the most probable value for the height of the burst-forming layer was 86 ± 2 km. The

question of a distribution of the bursts in height was examined by a method which was an elaboration of that given above, but no better measure of agreement with experiment was obtained than that yielded by the simple assumption of a thin layer of origin. It was therefore concluded that for oblique incidence and for radio frequencies in the range 20–45 Mc/sec. the bursts may be regarded as producing their scattered waves from within a very thin layer situated at a height of 86 km.

§ 6. ECHOING AREAS OF BURSTS

The range of echoing areas presented by ionospheric bursts was studied by the use of a simple arrangement of two parallel, horizontal half-wave dipoles each supported one quarter wavelength above the ground and operated as the transmitter and receiver aerials respectively. The radiation pattern of this system consisted of a single vertical lobe. In order to permit the estimation of actual echoing areas as distinct from relative values only, the system was standardized from observations made during test flights with an aircraft of known echoing area.

($P't$) records of burst occurrences were secured with this system and the range distribution histogram compiled as in figure 6. The mean rate of observed occurrences per unit area of the burst layer could then be calculated at ranges corresponding to the centres of the successive blocks of the histogram; assuming that the layer of occurrence is thin, for oblique incidence and at ranges much greater than 86 km. the volumes of successive zones of this layer will be proportional to the areas intercepted on the assumed surface of

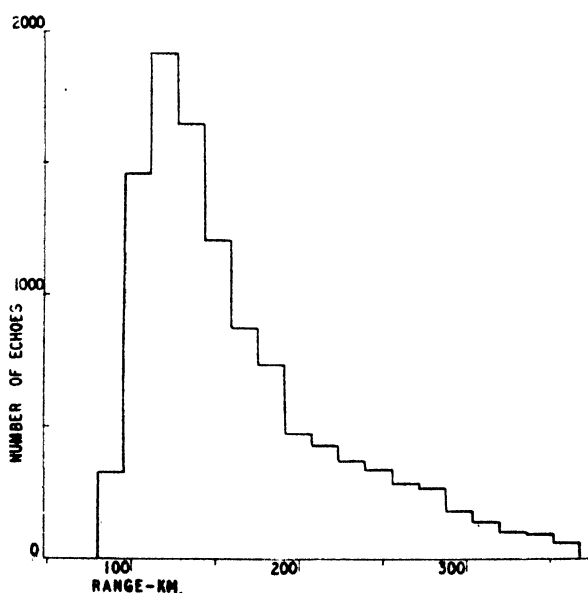


Figure 6. Range distribution.

occurrence situated at a height of 86 km. It is then sufficient to consider the superficial density of occurrence of the bursts on this equivalent surface.

Using the constants of the standardized aerial system the values were calculated of the smallest echoing areas that could be detected at points on the surface corresponding to the centres of the successive 16 km. zones. Let the minimum detectable echoing areas be A' and A'' for two zones for which the occurrence rate of observed echoes per unit area are N' and N'' respectively. If we assume that the distribution of bursts over the ionosphere is uniform then $(N' - N'')$ must be the rate of incidence per unit area of bursts having echoing areas between A and A'' .

The number of bursts per unit area of the ionosphere occurring in unit time which present echoing areas between A and $A+dA$ we have already defined as ndA , so that the number included in the interval A' to A'' will be $\int_{A'}^{A''} ndA$, i.e. $(N' - N'') = \Delta N = \int_{A'}^{A''} ndA$. Now this is the area under the (n, A) curve included between the A' and A'' ordinates, and since the mean ordinate over this interval, $\Delta N / (A' - A'')$, is readily calculable, the form of the (n, A) curve itself may be traced by plotting the equivalent rectangles included between ordinates erected at abscissae corresponding to the limiting echoing areas $A', A'',$ etc., of the successive zones. In figure 7 this method of plotting has been pursued and the straight line indicated leads immediately to the relation

$$ndA \propto A^{-3/2} dA. \quad \dots\dots(8)$$

It now follows that the number of bursts per unit area per unit time, N_A say, which present echoing areas greater than A will be given by

$$N_A = \int_A^\infty ndA \quad \text{i.e.} \quad N_A \propto 1/\sqrt{A}. \quad \dots\dots(9)$$

Separate confirmation of these results were obtained in the investigation on power effects previously mentioned.

From a set of observations made during January 1946 on a frequency of 22.69 Mc/sec. the constant of proportionality in equation (9) proved to be 643, and the units of area and time 1000 square miles and 50 hours respectively, which suggests that the total number of bursts per day created in the ionosphere with echoing areas greater than 10 m^2 is of the order 2×10^7 . This incidence rate may be compared with Skellett's assumption

(1935) of 10^9 sporadic meteors per day; it must be observed, however, that due to the comparatively low pulse recurrence frequency employed in the present work only echoes of duration greater than 0.3 sec. could be read from the films; also, no allowance has been made for a possible aspect effect such as was considered by Pierce (1938).

No upper limit has been observed for the echoing area which it is possible for a burst to present; thus in figure 5 for example the smallest echoing area which could be detected at the range corresponding to the first histogram minimum, i.e. 650 km., was of the order 10^7 m^2 , yet the figure shows that such echoes did occur.

A study of the rate of growth curves obtained from ciné observations on burst echoes shows that, in general, the echo grows rapidly in amplitude to a maximum during an interval of about $2/25$ sec. The subsequent decay process is much slower and is often closely exponential in form. Many echoes, on the other hand, do not show a smooth decay curve but one which oscillates in magnitude. Amplitude-time curves of this type have been obtained for a large number

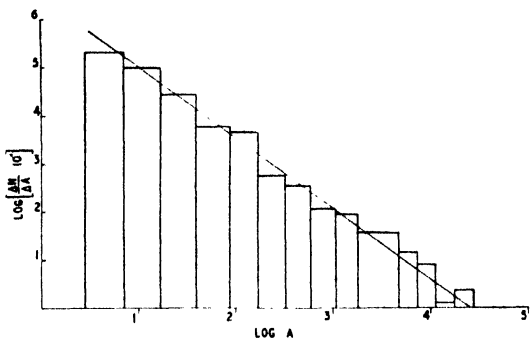


Figure 7.

of echoes originating from within the same small range interval. These curves show that the duration of a burst is not simply related to the maximum echoing area which it presents during its lifetime; they suggest that the processes of growth and decay of ionization must be subject to change from incident to incident, and that the manner of change in geometrical size of the ion cloud due to diffusion must also be variable.

§ 7. RELATIONSHIP BETWEEN APPARENT RATE OF OCCURRENCE OF BURSTS AND RADIO-FREQUENCY OF OBSERVING STATION

We have seen from figure 4 that the average level of activity during the eclipse experiments differed greatly from station to station, and it is of interest to examine whether this effect may be explained in terms of the known characteristics of the stations or whether the burst activity does indeed vary from locality to locality. In table 1 are set forth the average number of echoes per hour for each of the stations, together with the corresponding values of the integral $\int Q(R, h) dR$, for the appropriate range interval.

Table 1

Station	Frequency Mc/sec.	Echoes per hour between 120-220 miles = a	$\int Q(R, h) dR = b$	$m = a/b$
<i>Class A</i>				
Whalehead	25.10	802	109	7.4
Hillhead 1	22.69	607	80	7.6
Sango	25.10	256	33	7.7
Netherb'n	26.36	332	54	6.2
<i>Class B</i>				
Hillhead 2	42.27	236	109	2.2
Inverness	43.69	105	87	1.2
Tannach	42.27	88	94	0.9

The general conclusion drawn from this table is that the number of bursts observed by a station diminishes with increase of the radio frequency. We have found previously, however, that the echoing area distribution function expressed by equation (8) holds good for different radio-frequencies but with a different constant of proportionality. In order to interpret the table therefore it has appeared reasonable to modify equation (8) so as to include the effect of frequency explicitly and to write it in the form

$$n_r dA = A^{-3/2} c f(\nu) dA, \quad \dots (10)$$

where n_r is the occurrence rate of bursts per unit area which present echoing areas between A and $A + dA$ to a radio wave of frequency ν , $f(\nu)$ is some function of the frequency and c is a constant independent of the frequency.

The occurrence rate of bursts which present echoing areas greater than some value A now becomes $A^{-1} c f(\nu)$, and (7) takes the form:

Number of bursts in the range interval
 R_1 to R_2 observed to occur by a station operating on the radio-frequency ν :
 $\propto f(\nu) \int_{R_1}^{R_2} Q(R, h) dR. \dots (11)$

Following this simple treatment the unknown function $f(\nu)$ is clearly proportional to the value of the ratio m given in the final column of table 1, and it

would follow that stations operating on approximately the same radio-frequency should yield the same m . Examination of table 1 shows that the m -values for the Class A stations are substantially of the same magnitude and that $f(\nu)$ for these stations will also be approximately constant since the frequency range is small. This correspondence suggests that the differences in station activities are not due to local variations of burst activity but may be adequately explained in terms of the physical factors of position and type, of aerial, frequency and power radiated.

This question of local variation of burst activity was later examined in more detail by means of simultaneous observations made from two stations situated near Aberdeen and Weymouth respectively. It was not found that even this latitude separation produced any systematic activity differences.

The striking dependence of observed burst activity upon the radio-frequency employed is shown by comparing the results for the Class B stations with the figures for the A group. It is seen that doubling the radio-frequency produces a marked decrease in activity and a corresponding decrease in the m -values. The results are too rough to permit accurate deductions to be drawn and more work is clearly needed, but comparison of the average values of the two groups of stations show that $f(\nu)$ varies as $1/\nu^3$ and equation (10) becomes

$$n_e dA = \nu^{-3} A^{-3/2} c dA. \quad \dots (12)$$

It is clear that this apparent variation of burst activity with radio-frequency must be a consequence of the different echoing areas presented by a burst to radio waves of different frequencies and equation (12) leads immediately to the conclusion that the echoing area which a given burst presents to a wave of frequency ν varies inversely as ν^6 . Using this result it may now be deduced from equation (1) that the amplitude of the echo signal which a given burst returns to a station, other things being equal, will vary as the fourth power of the wavelength of the radio waves employed. It is interesting to note that this result is in agreement with Eckersley's long wavelength observations on scatter clouds (1932).

§ 8. GENERAL CONCLUSIONS WITH REFERENCE TO THE METEORIC THEORY

Reference has already been made to the development of the theory associating the occurrence of transient bursts of ionization in the ionosphere with the incidence of meteors, and it is of interest to examine whether the present experimental results are in agreement with this theory.

In a comprehensive account of meteor phenomena given by Hoffmeister (1937) a distinction is drawn between the shower type of meteor which originates within the solar system and pursues an elliptic orbit about the sun and the so-called sporadic or cosmic meteor that arrives from outer space. It is now known that a large majority of the meteors incident upon the earth are of the sporadic type and that definite diurnal and seasonal cycles are associated with their rate of occurrence at a given point upon the earth. On a typical night the rate of occurrence of meteors increases to a maximum in the early hours of the morning and then the activity falls off until dawn. Similarly the nocturnal activity follows a smooth seasonal cycle with its maximum in the autumn and minimum in the spring.

These variations of optical meteor activity have been explained in terms of the change of position of the observing station relative to the vector which defines

the earth's instantaneous orbital velocity. This vector is defined by its point of intersection with the celestial sphere, a point which is termed the meteoric apex. The maximum number of observable meteor impacts will occur when the hemisphere containing the station is the leading face of the earth in its orbital motion and when the meteoric apex lies on the station's meridian. This occurs at 0600 hours; also at 1800 hours there will be a minimum of impacts since the station then lies on the rearward face of the earth and the meteoric apex is at its lowest point. Other configurations will yield activities between these extreme values. Similarly, the total daily activity will change with the elevation of the diurnal path of the meteoric apex over the celestial sphere, and thus show a seasonal variation with maximum in the autumn.

It will be immediately noticed that the description we have given of the diurnal and seasonal changes of ionospheric burst activity closely parallels the above account of the time variation of sporadic meteor activity. A maximum of activity in the early morning and a minimum about 1800 hours are features of the diurnal cycle required by the above theory of meteors and these same features have been found characteristic of the daily curve of burst activity (figure 1). The form of the seasonal curve of burst activity given in figure 3 shows similar agreement with the above description of the seasonal variation of meteor activity. The amplitude of the early morning peak also shows a seasonal change in conformity with the changing angle of elevation of the meteoric apex, and in figure 1 it is seen that the autumn and winter amplitudes exceed those for winter and summer.

The late morning peak of the burst cycle, however, cannot be explained in terms of known characteristics of meteors, so that some other approach is necessary. Appleton and Naismith (1947) have recently found a correlation between the noon level of burst activity and the state of ionization of the lower E region, and they suggest that the susceptibility of the ionosphere to impact ionization by a meteoric particle probably varies with the existing intensity of ionization. The ionized state of the E region is subject to solar control in a manner which is fully understood, so that, following Appleton and Naismith, it now appears reasonable to attribute the second maximum of the daily cycle of burst activity to the changing sensitivity of the lower E region. From this it would immediately follow that the amplitude of this peak should vary seasonally and be a maximum at the summer solstice; figure 1 shows that this is the case.

It thus appears that the diurnal and seasonal cycles of burst activity may be explained in terms of two main factors, (a) the rate of incidence of meteors upon the area of the ionosphere visible from the station, (b) sensitivity of the burst-producing layer to impact ionization. Effect (a) is measured by the orientation of the earth's orbital motion vector relative to the station, while effect (b) depends upon the state of ionization of the E region which is controlled by the elevation of the sun. Both these effects produce an afternoon fall in activity and so must result in the pronounced minimum which we have observed to be so consistently present on our daily curves.

A further point of correspondence with the meteor theory is that of the heights of occurrence of the bursts. Olivier (1925) and Hoffmeister (1937) have collected the results of a number of observers on the mean heights of appearance and disappearance of meteors; the former value varies between 160 and 90 km., and disappearance occurs below 120 km. and mostly at about

80 km. In particular, Trowbridge (1907) has carefully studied meteor trains and finds that they tend to occur at a mean height of 87 km. These various figures may be compared with our result that the bursts appear to produce their scattered radiation from within a thin layer of the ionosphere at a height of 86 km.

We therefore conclude that the evidence of the present investigation supports the theory that a large proportion of ionospheric bursts may be attributed to impact ionization produced by sporadic meteors. The absence of an eclipse effect, moreover, is not at variance with this view, for the eclipse of the solar radiation would cause only a temporary and slight reduction in the sensitivity of the critical layer at a time coincident with the normal fall in activity, so that the detection of such a slight change would be difficult by the burst-counting method. The daily variation in the rate of incidence of meteors would also obscure the small effect that might be expected.

ACKNOWLEDGMENTS

This work was carried out under R.A.F. auspices by the permission of Air Ministry and the A.O.C. of No. 60 Group; it was initially prompted by Sir Edward V. Appleton, F.R.S., to whom our thanks are due for helpful discussion and advice. We would also like to acknowledge the valuable assistance received from many R.A.F. personnel and particularly from F/O. N. R. Benzie.

REFERENCES

- APPLETON, E. V., and NAISMITH, R., 1935, *Proc. Roy. Soc. A*, **150**, 685; 1940, *Proc. Phys. Soc.*, **52**, 402; 1946, *Nature, Lond.*, **158**, 936; 1947, *Proc. Phys. Soc.*, **59**, 461.
 APPLETON, E. V., NAISMITH, R., and INGRAM, L. J., 1937, *Philos. Trans. A*, **236**, 191.
 APPLETON, E. V., and PIDDINGTON, J. H., 1938, *Proc. Roy. Soc. A*, **164**, 467.
 ECKERSLEY, T. L., 1932, *J. Instn. Elect. Engrs.*, **71**, 405; 1937, *Nature, Lond.*, **140**, 846; 1940, *J. Instn. Elect. Engrs.*, **86**, 548.
 ECKERSLEY, T. L., and FARMER, F. T., 1945, *Proc. Roy. Soc. A*, **184**, 196.
 HEY, J. S., and STEWART, G. S., 1946, *Nature, Lond.*, **158**, 481.
 HOFFMEISTER, C., 1937, *Die Meteore* (Berlin: Springer).
 NAGAOKA, H., 1929, *Proc. Imp. Acad. Tokyo*, **6**, 5.
 OLIVIER, C. P., 1925, *Meteors* (Baltimore: Williams and Wilkins).
 PIERCE, J. A., 1938, *Proc. Inst. Radio Engrs.*, **26**, 892.
 SCHAFER, J. P., and GOODALL, W. M., 1932, *Proc. Inst. Radio Engrs.*, **20**, 1941.
 SKELLETT, A. M., 1932, *Proc. Inst. Radio Engrs.*, **20**, 1933; 1935, *Ibid.*, **23**, 132; 1938, *Nature, Lond.*, **141**, 472.
 TROWBRIDGE, C. C., 1907, *Astrophys. J.*, **26**, 95.

Measurements of the Interaction of Radio Waves in the Ionosphere

BY L. G. H. HUXLEY, H. G. FOSTER AND C. C. NEWTON

Department of Electrical Engineering, University of Birmingham

MS. received 28 September 1947

ABSTRACT. This paper describes measurements of radio wave interaction, in special test transmissions, between the B.B.C. transmitters at Droitwich and Lisnagarvey (N. Ireland). It is shown that the theoretical formula $M = M_0/[1 + (n/G\nu)^2]^{\frac{1}{2}}$ accurately describes the dependence of the interaction modulation on the modulation frequency $(n/2\pi)$. The electronic collisional frequency ν at the seat of interaction is deduced from $G\nu$ and the laboratory value of G . Investigations of the phase of the modulation are also described.

The paper contains a sketch of the theory of wave interaction. There is also an addendum dealing with the possible use of wave interaction as a tool in ionospheric research.

§ 1. INTRODUCTION

THIS paper describes the results of a number of observations of the phenomenon of non-resonant radio wave interaction in the ionosphere in special test transmissions. The equipment employed was similar to, but an improvement on, that used by us in earlier observations on gyro-interaction that have already been reported (Huxley, Foster and Newton 1947). The test transmissions which concern us in this paper were arranged by Mr. J. A. Ratcliffe of the Cavendish Laboratory in consultation with the Engineering Research Department of the British Broadcasting Corporation, and through their courtesy we were informed of these transmissions. Mr. Ratcliffe also informed us of his method for roughly locating the region of interaction described in § 5 below.

A special effort was made in our observations to obtain accurate measurements of the depth of interaction modulation as a function of audio-frequency.

The transmitters used in the tests were Droitwich (170 kw., 200 kc/s., 80% modulation) and Lisnagarvey, N. Ireland (60 kw., 1050-kc/s., plain carrier), although Droitwich was replaced by Ottringham, near Hull (170 kw., 167 kc/s.), on a few occasions. The point of observation was in Selly Oak, Birmingham. Their relative positions are shown in figure 1.

Before proceeding to a description of the experimental results, a brief account of the phenomenon of wave interaction and the elements of the theory are given in order that the significance of the observations may be the better appreciated.

§ 2. DESCRIPTION OF THE PHENOMENON AND NOMENCLATURE

Consider two radio transmitters; the one, which we shall call the interfering transmitter, radiates at high power a modulated carrier, whereas the carrier of the other, called the wanted transmitter, is unmodulated. An observer, so placed that he can receive the components of the transmission from the wanted station that have passed through the E region of the ionosphere in the vicinity of the interfering station, finds that, in general, the modulation of the interfering transmission is impressed upon the wanted transmission in the course of its passage through the E region. This phenomenon is called Wave Interaction (previously the Luxembourg Effect). The phenomenon was first reported by Tellegen (1933) and the first satisfactory attempt at a theoretical explanation was made by Bailey and Martyn (1934). Later Bailey (1937) revised the original theory and gave a more rigorous and complete treatment. Systematic investigations of the phenomenon were made by Grosskopf (1938).

§ 3. SIMPLIFIED THEORY OF WAVE INTERACTION

The basic ideas of the theory given by Bailey and Martyn and later elaborated by Bailey are the following. The absorption coefficient k_0 of the ionosphere for the wanted wave whose frequency is $p/2\pi$ at a place where the collisional frequency is ν is known to be (neglecting the earth's magnetic field)

$$k_0 = \frac{2\pi e^2 c N}{m} \left(\frac{\nu}{\nu^2 + p^2} \right) \text{cm}^{-1}, \quad \dots\dots (1)$$

where e is the electronic charge in E.M.U., m the mass of the electron, $c = 3 \times 10^{10}$ and N is the electronic concentration in cm^{-3} . The absorption coefficient is thus a function of the electronic collisional frequency and the frequency of the wanted wave.

The E region in the vicinity of the interfering station (frequency $p_1/2\pi$) is irradiated by the electric field of the latter,

$$E_1(1 + D \cos nt) \cos p_1 t, \quad \dots\dots(2)$$

where D is the modulation depth and $n/2\pi$ the modulation frequency. Suppose first that the modulation D is zero. The field $E_1 \cos p_1 t$ then supplies power W to an electron at a specified region in the ionosphere. When W is zero, the kinetic energy of agitation of an electron is the same as the thermal energy of agitation Q_t of the molecules of the gas. When, however, power W is supplied to an electron at a steady rate, a new condition of dynamical equilibrium results in which the mean kinetic energy $\bar{Q}$ of the electron exceeds that of the gas molecules, so that we may write $\bar{Q} = k_T Q_t$, where k_T is Townsend's energy factor.

In this condition, let the collisional frequency of the electron be $\bar{\nu}$ and $\gamma\bar{Q}$ the mean energy lost per collision with gas molecules. Then

$$\gamma\bar{\nu}\bar{Q} = W. \quad \dots\dots(3)$$

When the collisions between electrons and gas molecules resemble those between smooth rigid spheres, it is known that with a Maxwellian distribution of velocities

$$\gamma = 2.66(\mathbf{m}/M')(1 - 1/k_T) = 2.66(\mathbf{m}/M')(1 - \nu_t^2/\bar{\nu}^2),$$

where M' is the mass of a molecule of the gas and ν_t is the collisional frequency of the electron when in thermal equilibrium with the gas ($W=0$).

Bailey has supposed that a similar formula is applicable to an electronic motion in a diatomic gas when $2.66\mathbf{m}/M'$ is replaced by a factor G which is characteristic of the gas. Thus,

$$\gamma = G(1 - 1/k_T) = G(1 - \nu_t^2/\bar{\nu}^2). \quad \dots\dots(4)$$

In general, G greatly exceeds $2.66\mathbf{m}/M'$, as for instance in air, where $2.66\mathbf{m}/M' = 5 \times 10^{-5}$ but $G \simeq 10^{-3}$.

When the coefficient of modulation D in equation (2) is zero, suppose E_1 to supply power to an electron at the mean rate

$$W = AE_1^2, \quad \dots\dots(5)$$

where A depends on the frequency $p_1/2\pi$, the collisional frequency $\bar{\nu}$ and $\mathbf{e}^2/\mathbf{m}$ (Appendix).

When D is not zero the power supplied to an electron becomes time-dependent as follows:

$$W(t) = W(1 + D \cos nt)^2 = \bar{W} + 2W(D \cos nt + \frac{1}{4}D^2 \cos 2nt), \quad \dots\dots(6)$$

where

$$\bar{W} = (1 + D^2/2)W. \quad \dots\dots(7)$$

The energy of agitation Q now becomes a function of the time which we determine.

Let $\bar{Q}$ be the mean value of Q when power $\bar{W}$ is communicated to each electron, and Q_t its value when $\bar{W}=0$ (thermal equilibrium). It follows from (4) and (6) with $k_T = \bar{Q}/Q_t = (\bar{\nu}/\nu_t)^2$, that

$$dQ/dt + (G\bar{\nu})(Q - Q_t) = W(t). \quad \dots\dots(8)$$

Similarly,

$$(G\bar{\nu})(\bar{Q} - Q_t) = \bar{W} = W(1 + D^2/2). \quad \dots\dots(9)$$

Hence

$$dQ/dt + (G\bar{\nu})(Q - \bar{Q}) = 2W[D \cos nt + \frac{1}{4}D^2 \cos 2nt], \quad \dots\dots(10)$$

$$\text{whence } Q = \bar{Q} \left[1 + \frac{2W}{\bar{Q}(G\bar{\nu})} \left\{ \frac{D \cos (nt - \phi)}{(1 + (n/G\bar{\nu})^2)^{1/2}} + \frac{D^2 \cos (2nt - \phi')}{4(1 + (2n/G\bar{\nu})^2)^{1/2}} \right\} \right], \quad \dots\dots(11)$$

where

$$\tan \phi = (n/G\bar{\nu}); \quad \tan \phi' = (2n/G\bar{\nu}). \quad \dots\dots(12)$$

Thus the agitational energy Q fluctuates about its mean value $\bar{Q}$ at a frequency $(n/2\pi)$. Since the collisional frequencies ν and $\bar{\nu}$ corresponding to Q and $\bar{Q}$ satisfy the relation $(\nu/\bar{\nu})^2 = (Q/\bar{Q})$, and since in practice the coefficient $2W/[\bar{Q}(G\bar{\nu})] \ll 1$ in (11), it follows that ν is modulated as follows:

$$\nu = \bar{\nu} \left[1 + \frac{W}{\bar{Q}(G\bar{\nu})} \left\{ \frac{D \cos(nt - \phi)}{(1 + (n/G\bar{\nu})^2)^{1/2}} + \frac{D^2 \cos(2nt - \phi')}{4(1 + (2n/G\bar{\nu})^2)^{1/2}} \right\} \right]. \quad \dots\dots (13)$$

Thus the ratio of the amplitude of the second harmonic to that of the fundamental in the modulation of Q and ν is

$$\frac{1}{4} D \{n^2 + (G\bar{\nu})^2\} / [4n^2 + (G\bar{\nu})^2]. \quad \dots\dots (14)$$

It is the custom to neglect the second harmonic but its existence in the correct proportion required by (14) has been established in measurements of wave interaction (Huxley, Foster and Newton 1947).

In what follows we consider the fundamental term alone and write (13) $\nu = \bar{\nu} + \nu_n$ with

$$\nu_n = \frac{\bar{\nu} W D}{\bar{Q}(G\bar{\nu})} \frac{\cos(nt - \phi)}{[1 + (n/G\bar{\nu})^2]^{1/2}}. \quad \dots\dots (15)$$

It follows from (9) and (14) that the coefficient $WD/\bar{Q}(G\bar{\nu})$ in (13) may be written in the alternative forms

$$\frac{WD}{\bar{Q}(G\bar{\nu})} = \frac{WD}{\bar{W} + Q_t(G\bar{\nu})} = \frac{D}{[(1 + D^2/2) + Q_t(G\bar{\nu})/W]}, \quad \dots\dots (16)$$

and
$$WD/[\bar{Q}(G\bar{\nu})] = D(k_T - 1)/(1 + D^2/2)k_T. \quad \dots\dots (17)$$

When the wanted wave traverses a region in which the collisional frequency is modulated as in (15) by the field of the interfering station, the absorption coefficient (equation (1)) also becomes modulated as follows:

$$k = k_0 + \nu_n dk_0/d\nu. \quad \dots\dots (18)$$

Suppose the modulation of k to occur within a region which is traversed by the wanted wave along a ray of length s . The emergent amplitude of the wanted wave in terms of its amplitude on entering the region is

$$E = E_0 \exp \left(- \int_0^s k ds \right) = E_0 \exp \left(- \int_0^s k_0 ds \right) \cdot \exp \left(- \int_0^s \nu_n (dk_0/d\nu) ds \right). \quad \dots\dots (19)$$

Since the total attenuation of the wave is in practice not excessive, the exponent $-\int_0^s k_0 ds$ is of the order of magnitude of unity or less. The increment in the absorption is still less and we may write $\int_0^s \nu_n (dk_0/d\nu) ds \ll 1$; consequently (19) may be written,

$$E = E_0 \exp \left(- \int_0^s k_0 ds \right) \left[1 - \int_0^s \nu_n (dk_0/d\nu) ds \right]. \quad \dots\dots (20)$$

Thus, the depth of the impressed modulation, expressed as a fraction of the received amplitude of the wanted wave is

$$M = - \int_0^s \nu_n (dk_0/d\nu) ds. \quad \dots\dots (21)$$

Suppose for simplicity that k_0 and $\bar{\nu}$ are constant along the path of integration through the region of interaction. This implies either that the path of integration lies at the top of the trajectory of the ray, or that, if the interaction region lies on the incident or reflected ray, it is effectively localized. The expression for M then simplifies (from (16)) to

$$M = \frac{D\bar{\nu}(dk_0/d\nu)s}{[1 + D^2/2 + (G\bar{\nu})Q_t/W]} \frac{\cos(nt - \phi)}{[1 + (n/G\bar{\nu})^2]^{\frac{1}{2}}}. \quad \dots\dots (22)$$

Which alternative is the more appropriate is to be decided from experience. In any case, it is supposed that where interaction occurs $p^2 \gg \bar{\nu}^2$, so that according to (1) the absorption coefficient is proportional to $\bar{\nu}$. The total absorption of the wanted wave in nepers is $\int_0^s k_0 ds$, which is, therefore, of the form $C\bar{\nu}$ where C is determined by medium and the path. The term $s(dk_0/d\nu)\bar{\nu}$ in (22) may be replaced by $(C\bar{\nu}s)$, which can be estimated from the reflection coefficient of the E region for the wanted wave.

The expression (22) for M shows that in this case the dependence of M on the modulation frequency is relatively simple and of the form

$$M = M_0/[1 + (n/G\bar{\nu})^2]^{\frac{1}{2}}, \quad \dots\dots (23)$$

where, from (22), $M_0 = DCs\bar{\nu}/[(1 + D^2/2) + Q_t(G\bar{\nu}/W)]$.

Further, the interaction modulation lags in phase with respect to the interfering field E_1 at the region of modulation by the angle

$$\phi = \tan^{-1}(n/G\bar{\nu}). \quad \dots\dots (24)$$

In the measurements described below special attention was paid to the investigation of M as a function of n and it was found that formula (23) describes accurately what is observed when $n/2\pi$ does not exceed 750 c/s. From such measurements the quantity $G\bar{\nu}$ may be deduced in each instance.

In practice, the term $(G\bar{\nu})Q_t/W \gg (1 + D^2/2)$ in (22) so that M_0 is effectively equal to $WDCs\bar{\nu}/Q_t(G\bar{\nu})$.

§ 4. MEASUREMENT OF INTERACTION MODULATION AS A FUNCTION OF MODULATION FREQUENCY

(i) Calibration of equipment

Because the depths of interaction modulation are relatively small in practice, it was necessary to use measuring equipment capable of giving reliable results down to modulation depths of about 0.1% on a carrier received from a relatively distant (300–400 km.) medium wave station. It was decided therefore to use not an oscilloscope but a specially calibrated receiver and to obtain the carrier voltage and the audio-frequency voltage at the detector from meter readings, from which the modulation depth could be obtained from calibration curves. The receiver was a modified commercial communications receiver which covered the required range of wavelengths. The detector stage was modified to feed directly into a valve voltmeter circuit which was built into the receiver chassis, and which was calibrated to provide the carrier level at the detector in the form of a meter reading.

The audio frequency voltages from the detector were fed, through suitable filter circuits, to a cathode follower, the output of which was connected directly to a Marconi wave analyser which was available in the laboratory. The advantage of a wave analyser lies in its narrow bandwidth of only 4 c/s. which enables the

wanted modulation tone from the detector to be measured relatively free from background noise and interference.

Through a laboratory calibration it was possible to use the direct readings of the wave analyser in combination with those of the carrier-level meter mentioned above to obtain the required modulation depth immediately. The wave analyser had, in fact, been originally used to measure the second harmonic of the interaction modulation shown in equation (13) as mentioned in the paragraph that follows it.

The calibration of the system to read modulation depths direct was achieved by applying to the input, a radio frequency carrier whose frequency was near that of the wanted transmission. This carrier was modulated by a series of pure tones covering the range of frequencies of 50 to 2000 c/s. The modulation depths, for the purposes of calibration were set in turn at $2\frac{1}{2}\%$, 5% and 10%. Later this calibration was found to agree with a calibration tone of known depth specially imposed upon the carrier of the wanted transmission radiated by Lisnagarvey (and Westerglen on another occasion).

The laboratory calibration unit comprised a crystal oscillator driving a power amplifier which was modulated by the appropriate tone.

By extrapolation of the calibration curves which are straight lines, modulation depths down to 0.1% could be confidently assessed.

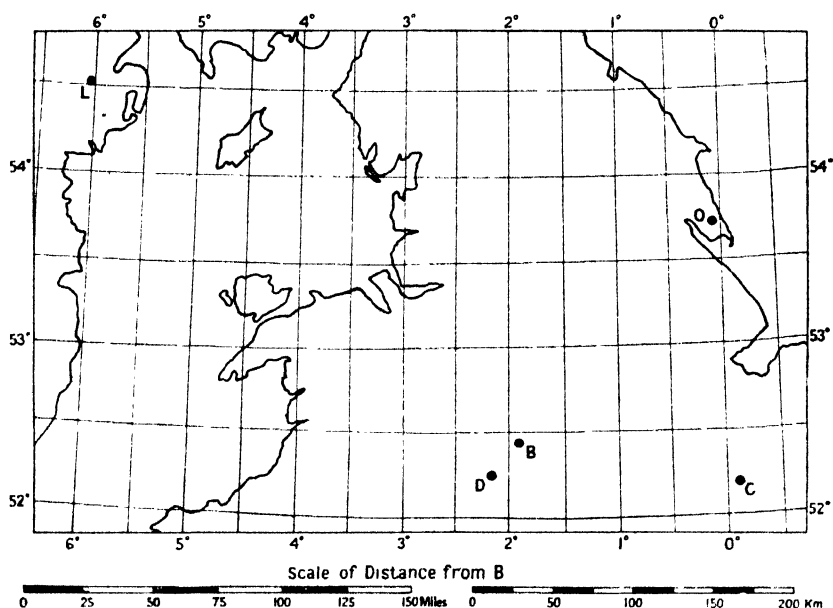


Figure 1. Gnomonic projection centred on B, showing stations and point of observation.

(ii) *Experimental results*

The transmitters concerned are shown in figure 1, in which D, L and O represent respectively Droitwich, Lisnagarvey and Ottringham. The point of observation was in Birmingham and is indicated by B. The distance LB is 340 km., whereas DB is 20 km. and the angle LBD is approximately a right angle.

The transmitter at D radiated 170 kw. in the form of a carrier on 200 kc/s. modulated by a series of pure tones to a depth of 80%, while at the same time the transmitter L radiated 60 kw. as a plain carrier on a frequency of 1050 kc/s.

The times of observation were between the hours of 2230–0300 G.M.T. on the nights of June 15/16 and 16/17, July 3/4, 4/5, 5/6 and 6/7, and August 6/7, 7/8 and 8/9, 1947.

At intervals during the night, pulse transmissions were sent out by Lisnagarvey which showed that the ground wave received in Birmingham was relatively weak, and that the predominant wave received during the early night was a single reflection from the E region, often accompanied later by other forms of reflected waves from the E or F regions with relative strengths varying continually.

In the test runs, in which M was determined as a function of $(n/2\pi)$ the modulation frequency, Droitwich was modulated in turn, for two minutes each, at the frequencies 50, 75, 100, 150, 200, 300, 400, 500, 750, 1000, 1250, 1750 and 2000 c/s.

The modulation depths were found from simultaneous readings of the carrier level meter and the wave analyser as described in the previous section, using A.v.C. to hold the carrier level constant at the detector. It was confirmed that the use of A.v.C. did not affect the measured modulation depth, except to improve the accuracy.

Two examples of the results of such measurements are shown in the curves of figures 2(a) and (b). They refer to the times stated on the figures.

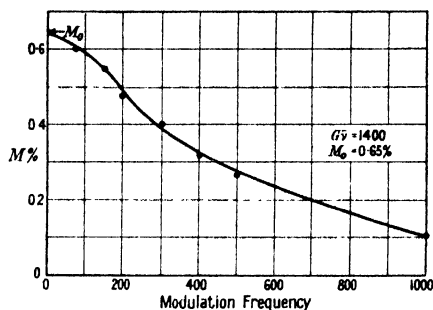


Figure 2(a). Interaction modulation depth *vs* modulation frequency —Droitwich—Lisnagarvey 0230–0300 G.M.T. 16/6/47.

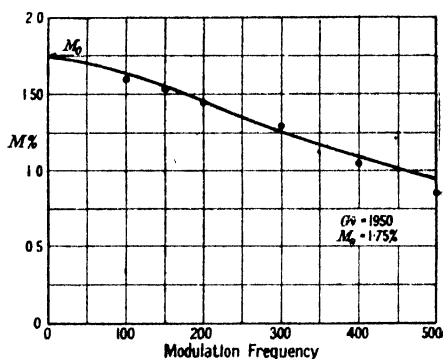


Figure 2(b). Interaction modulation depth *vs* modulation frequency —Droitwich—Lisnagarvey 2330–2400 G.M.T. 3/7/47.

If the theoretical formula (23),

$$M = M_0 / [1 + (n/G\bar{v})^2]^{\frac{1}{2}} \quad \dots\dots (25)$$

actually describes the dependence of M on n , then M and n should also be related as follows:—

$$n^2 = M_0^2 (G\bar{v}/M)^2 - (G\bar{v})^2, \quad \dots\dots (26)$$

in which M_0 is the value of M as $n \rightarrow 0$.

Thus, the curve which represents n^2 as a function of $1/M^2$ is a straight line whose intercept on the n^2 axis is $(G\bar{v})^2$. In figures 3(a) and (b) the experimental data exhibited in figures 2(a) and (b) are plotted with n^2 as a function of $1/M^2$. It can be seen that for frequencies $n/2\pi$ from 50 to 500 c/s. a linear relation subsists; consequently, at the lower frequencies $n/2\pi$, where the measurements are more accurate, equation (25), does in fact closely describe the experimental facts.

Evidently the quantity $G\bar{v}$ may be found from the intercept on the n^2 axis of the straight line through the experimental points. Further, M_0 is found from the intercept on the $1/M^2$ axis.

It is therefore necessary only to give the zero frequency modulation M_0 and the value of $G\bar{\nu}$ in order to specify completely the modulation M at any frequency $n/2\pi$.

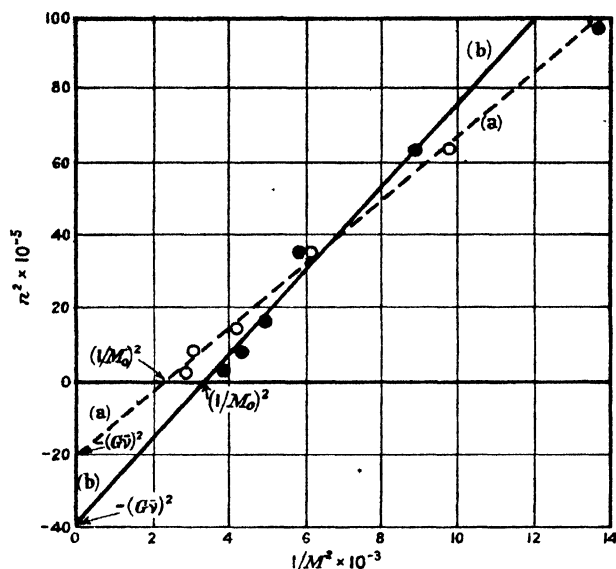


Figure 3. Points in figure 2 replotted to exhibit $1/M^2$ as a function of $n^2 = (3\pi f)^2$.

In what follows the modulation is specified in this manner.

Tables 1 and 2 summarize the experimental results on the dependence of M on n . The values of $G\bar{\nu}$ ranged from 1100 to more than 2000 in these tests.

Table 1. Observed depths on interaction modulation. Wanted station—Lisnagarvey (1050 kc/s.)

Interfering transmitter—Droitwich (200 kc/s.; 170 kw.) 80%.							
Date	Time G.M.T.	M_0 (%)	$G\bar{\nu}$	Date	Time G.M.T.	M_0 (%)	$G\bar{\nu}$
June 16	0230-0300	0.6	1400	July 5	2230-2300	1.63	2320
„ 16	2300-2330	1.75	1480	„ 5	2330-2400	2.5	—
„ 17	0230-0300	0.9	irregular	„ 6	2230-2300	1.2	2860
July 3	2330-2400	1.75	1950	„ 6	2330-2400	1.4	1480
„ 4	2230-2300	1.75	1870	„ 7	0130-0200	2.1	1480
„ 4	2330-2400	1.85	1520				

Interfering transmitter—Ottringham (167 kc/s.; 200 kw.) 80%.

Date	Time G.M.T.	M_0 (%)	$G\bar{\nu}$
June 16	0200-0230	0.58	1550
„ 17	0200-0230	0.45	1480

Table 2. Observed depths of interaction modulation. Wanted station—Lisnagarvey (1050 kc/s.)

Interfering transmitter—Droitwich (200 kc/s.; 170 kw.).											
Date	Time	G.M.T.	M_0 (%)	$G\bar{\nu}$	h (km.)	Date	Time	G.M.T.	M_0 (%)	$G\bar{\nu}$	h (km.)
Aug. 6	2300–2328		2.6	1550	75	Aug. 7	2300–2328		2.6	1340	95
„ 6	2340–2348		2.6	1100	68	„ 7	2340–2348		2.1	—	86
„ 7	0010–0018		1.55	1050	70	„ 8	0010–0018		1.9	1720	86
„ 7	0040–0048		2.2	1570	88	„ 8	0040–0048		2.25	1870	86
„ 7	0130–0158		1.9	1450	88	„ 8	0130–0158		2.1	1220	97
„ 7	0210–0218		—	—	83	„ 8	0210–0218		1.5	1790	86
„ 7	0240–0248		2.0	1950	83	„ 8	0240–0248		1.5	—	86
						„ 9	0100–0128		0.77	1140	82

h = height of interaction.

It was found in an isolated observation of wave interaction between Westerglen (767 kc/s.) near Falkirk and Ottringham (167 kc/s.-170 kw.) on 8th May (0030-0100) that $M_0 = 4\%$ and $G\bar{\nu} = 1800$. Thus, $G\bar{\nu}$ is of the order, in these experiments, of 1.5×10^3 . If it be assumed that the temperature and composition of the air at the region of interaction are approximately the same as those at the ground, then, as Bailey has done, we may suppose that $G = 2.6 \times 10^{-3}$, deduced from the experiments of Townsend and Tizard on the diffusion of slow electrons in air. Then from the experimental value of $G\bar{\nu}$ we find that $\bar{\nu}$ is approximately 6×10^5 , and exceeds ν_t by a few per cent.

Recent experiments to determine G by Mr. Zaazou and one of us in this department have shown that its value for air is 1.3×10^{-3} and not 2.6×10^{-3} as supposed hitherto. Thus $\bar{\nu}$ becomes 1.2×10^6 . Further, if oxygen is present as atomic oxygen at the interaction region, then G should be somewhat smaller than the laboratory value.

§ 5. THE PHASE OF THE INTERACTION MODULATION

It was pointed out by Ratcliffe that an examination of the phase of the interaction modulation should provide a further test of the theory. According to equations (12), the phase of the modulation of the collision frequency lags on that of the interfering field E_1 at the place of interaction by an angle $\phi = \tan^{-1}(n/G\bar{\nu})$. It follows therefore, from (22) that the phase of the interaction modulation, with that of interfering field as the standard of reference as above, lags by $\phi = \tan^{-1}(n/G\nu_0)$ or $\tan^{-1}(n/G\nu_0) \pm 180^\circ$ according as $(dk_0/d\nu)$ in (21) has a negative or a positive sign. Since, from (1), k_0 is a maximum when $p = \nu$, and $(dk/d\nu)$ is negative when $\nu > p$, but positive when $\nu < p$, it follows that the very low frequency modulation $M_0(n \rightarrow 0)$ should vibrate in phase with the interfering modulation if interaction occurs below the height where $\nu = p$, but should vibrate in anti-phase if injected in a region at a greater height.

It was found that the zero frequency modulation was anti-phased with respect to that of the interfering field. It may therefore be concluded that interaction takes place above the level at which $\nu = p$ (where $p/2\pi$ is the frequency of the wanted wave).

In figure 4, D represents the interfering transmitter (Droitwich), B the point of observation (Selly Oak, Birmingham) and I the centre of a region where interaction occurs. Thus, IB is a ray from the wanted station (Lisnagarvey) that reaches B. Let $DB = l_1$ (20 km.) and put $l_2 = DI + IB$. Take as reference phase for the modulation tones that of the interfering modulation at D, its source. The phase lag ψ_1 of this modulation when received at B on the carrier of the interfering station is

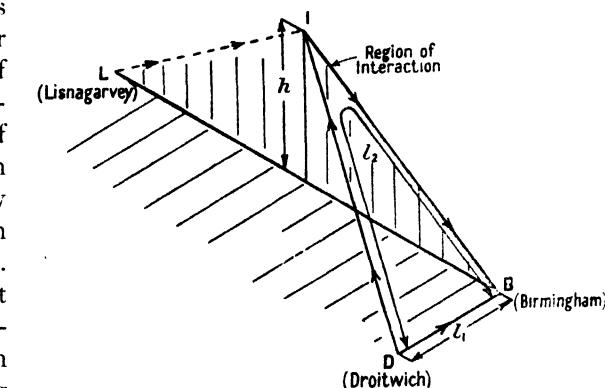


Figure 4. Relative positions of the wanted and interfering stations and of the reception point.

is

$$\psi_1 = 360fl_1/v \text{ degrees,} \quad \dots\dots (27)$$

where the modulation frequency $f = n/2\pi$ c/s. and $v = 3 \times 10^5$ km/sec.

Similarly, the phase of the interaction modulation generated at I, when received at B on the carrier of the wanted wave, lags on the reference phase at D by

$$\psi_2 = 360fl_2/v + \tan^{-1}(2\pi f/G\bar{v}) \pm 180 \text{ degrees.} \quad \dots\dots(28)$$

Thus, if the audio outputs of identical receivers are compared on a double-beam oscilloscope, the relative phase lag at B of the interaction modulation with respect to the interfering modulation, which is

$$(\psi_2 - \psi_1) = 360f(l_2 - l_1)/v + \tan^{-1}(2\pi f/G\bar{v}) \pm 180 \text{ degrees,} \quad \dots\dots(29)$$

may be measured. If the receivers are not identical the phase shift introduced by each at any frequency f is easily determined and the difference used to correct the value of $(\psi_2 - \psi_1)$ given by the oscilloscope.

Let ψ be the frequency-dependent part of (29), that is

$$\psi = 360f(l_2 - l_1)/v + \tan^{-1}(2\pi f/G\bar{v}_0). \quad \dots\dots(30)$$

It follows that $\psi = 0$ when $f = 0$, and that

$$\psi \rightarrow 360f(l_2 - l_1)/v + 90^\circ, \text{ as } f \rightarrow \infty.$$

Thus the curve representing ψ as a function of f passes through the origin and possesses an asymptote whose slope is

$$f \rightarrow \infty (d\psi/df) = 360(l_2 - l_1)/v, \quad \dots\dots(31)$$

and whose intercept with the ψ -axis is 90° .

When the measured values of ψ are plotted against f , the distance l_2 may be obtained from the slope of the asymptote as follows:

$$l_2 = l_1 + \frac{1}{12}(d\psi/df)10^4. \quad \dots\dots(32)$$

It follows that if the elevation of the downcoming ray IB and the angle LBD are known, then the point I may be determined. This method of fixing I from (30) and (32) was pointed out to us by J. A. Ratcliffe.

Figure 5 shows that a typical experimental curve representing ψ as a function of f conforms closely to what is to be expected from (30). On the assumption that the region of interaction I lies on a ray reflected at a height of 100 km., it was found that in most cases the apparent interaction point obtained from (32) was located at a height of (86 ± 5) km. The method therefore provides a rough estimate only of the height.

The most stable readings of amplitude and phase are obtained when one of the waves reflected from the

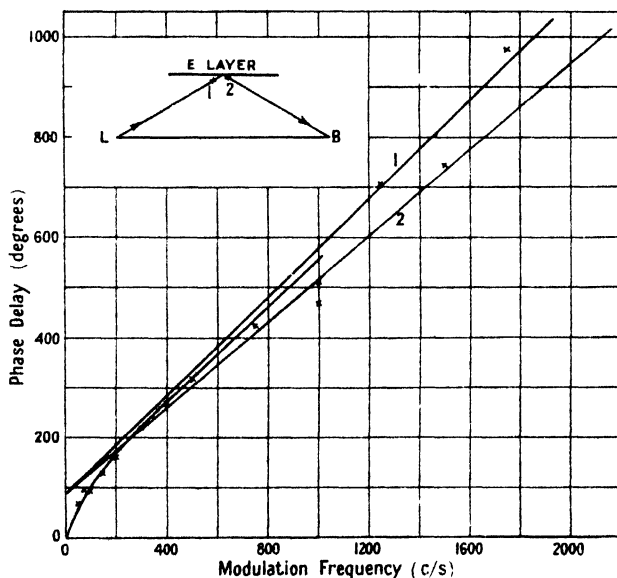


Figure 5.

Note. Inset: Point 1 obtained from asymptote 1.
Point 2 obtained from asymptote 2.

ionosphere is predominant. In these tests the predominant wave was the wave undergoing a single reflection at the E region as was shown by pulse transmissions from Lisnagarvey. However, except during the early evening, other waves reflected from the F or E regions appeared in varying strengths, and the interaction modulation on these components was able, on occasions, to shift the phase in an irregular fashion and to alter the amplitude of the interaction as measured at the point of observation.

It was noted that when A.V.C. was employed in measurements of modulation depth, the audio-frequency amplitude showed sudden departures from a steady value, although the carrier was held fixed. The steady level was assumed to correspond to the single E wave and the fluctuations to the presence of other sky waves.

§ 6. REMARK ON THE MAGNITUDE OF THE INTERACTION

The maximum depths of interaction modulation M_0 measured in these tests were small, as can be seen from tables 1 and 2, and we believe that this relatively small value is to be attributed to the relatively high frequency of the wanted station (1050 kc/s.).

It is of interest to compare the values of M_0 listed in tables 1 and 2 with those obtained in other transmissions.

Thus, van der Pol and van der Mark found that the interaction M_0 impressed on Beromunster (556 kc/s.) by Luxembourg (150 kw.) was about 8%, and as mentioned below table 1, we observed that Ottringham (100 kw.) impressed upon Westerglen (767 kc/s.) a modulation M_0 of about 4%.

In these tests the modulations M_0 impressed upon Lisnagarvey (1050 kc/s.) by Droitwich (170 kw.) were less than 2%. Thus, M_0 appears to diminish as p increases.

ACKNOWLEDGMENT

The authors wish to express their appreciation of the co-operation and interest shown by Mr. H. L. Kirke and Mr. R. A. Rowden of the Engineering Research Section of the British Broadcasting Corporation in providing them with schedules of the transmissions.

ADDENDUM

Suggested measurements of wave interaction

By L. G. H. HUXLEY

It is explained, in the above paper, how a determination of the amplitude and phase of the impressed modulation M as a function of modulation frequency $n/2\pi$ leads to value of the product $G\bar{v}$. It is natural to enquire whether the phenomenon of interaction of radio waves is able to provide estimates of the value of other molecular and electronic parameters of the ionosphere. To this end, consider the general formula (22) for the impressed modulation. According to this formula, and what immediately follows it, the amplitude of the zero frequency modulation M_0 is

$$M_0 = \frac{DC\bar{v}s}{(1 + D^2/2 + (G\bar{v})Q_t/W)}. \quad \dots\dots (A\ 1)$$

The mean carrier power W supplied by the interfering field E_1 to an electron is proportional to the radiated power P of the interfering transmitter, so that in

general M_0 is not proportional to the radiated power P , but becomes so in effect when the term $(G\bar{\nu})Q_t/W \gg 1$.

We therefore first consider the order of magnitude of $(G\bar{\nu})Q_t/W$ in the practical case where $P=200$ kw. and the frequency $p_1/2\pi=2 \times 10^5$. We may consider, with good approximation to the truth, that the aerial with its image radiates as a Hertz doublet.

The amplitude E_1 of the field produced at a point at height h in the E layer with an angle of elevation α with respect to the transmitter is given by

$$E_1^2 = 4.5 \times 10^{-6} (P \sin 2\alpha) / h^2 \text{ (volts/cm)}^2, \quad \dots\dots (A\ 2)$$

where P is in kw., h in km. and E_1 in v/cm.

It follows that the amplitude (peak) of the field E_1 when $h=100$ km., and $P=200$ kw. is $2.12\sqrt{2} \times 10^{-4} \sin 2\alpha$ v/cm.

To estimate W , assume a collisional frequency $\bar{\nu}=10^6$. We may suppose therefore that $p_1^2 = (4\pi \times 10^5)^2 \simeq \bar{\nu}^2$, and proceed to estimate $W = AE_1^2$, with

$$A \simeq \left\{ \frac{1}{2} e^2 \bar{\nu} / m (p_1^2 + \bar{\nu}^2) \right\} \cos^2 \theta$$

We find $W = 1.8 \times 10^{-11} \cos^2 \theta \sin 2\alpha$ erg/sec. If the temperature of the region where interaction occurs lies between 200° and 300° K., then Q_t is approximately 5×10^{-14} erg. Since $G\bar{\nu} \simeq 1.5 \times 10^3$, we find

$$(G\bar{\nu})Q_t/W \simeq 4.2 / \cos^2 \theta \sin 2\alpha.$$

The maximum modulation is obtained when the point of observation is so chosen, if possible, that θ , the angle between E_1 and the earth's magnetic field, is zero and α is 45° . The term $(G\bar{\nu})Q_t/W$ in (A 1) is then approximately $(1 + D^2/2)$ and the impressed modulation M_0 would be less at $P=200$ kw. than would be predicted from its value when P is small, on the assumption that M_0 is proportional to P .

If, however, the maximum power P were notably less than 200 kw. or the angles θ and α appreciably different from zero and 45° respectively, then $(G\bar{\nu})Q_t/W$ at maximum power would certainly exceed $(1 + D^2/2)$ by a factor of eight or more. An observer, unless his measurements of modulation M_0 were very accurate, would judge M_0 to be proportional to P up to the maximum value of P .

If, on the other hand, a very powerful transmitter capable of radiating 300 kw. or more, or a strong gyro-transmitter, were employed to produce the interfering field, then measurements of M_0 by the sensitive method described in §4.1 should reveal the lack of proportionality between M_0 and P under optimum geometrical conditions.

Such measurements should also provide additional information about the physical properties of the E layer, as may be appreciated from the following considerations.

For equation (A 2) write $E_1^2 = KP$, then, $W = AE_1^2 = AKP = bP$, where

$$b = 10^{-6} \frac{e^2}{m} \cdot \frac{\bar{\nu}}{(p_1^2 + \bar{\nu}^2)} \cdot \frac{\cos^2 \theta \sin 2\alpha}{h^2}. \quad \dots\dots (A\ 3)$$

Write (A 1) in the form

$$M_0 = a / [1 + D^2/2 + (G\bar{\nu})Q_t/bP], \quad \dots\dots (A\ 4)$$

where, from (1),

$$a = DC\bar{\nu}s. \quad \dots\dots (A\ 5)$$

Expression (A 4) is equivalent to

$$1/M_0 = (1 + D^2/2)/a + (G\bar{\nu})Q_t/abP. \quad \dots\dots (A\ 6)$$

If, therefore, the experimental values of $1/M_0$ were plotted against $1/P$, the points should fall near a line whose slope is $(G\bar{\nu})Q_t/ab$ and whose intercept on the $1/M_0$ axis is $(1 + D^2/2)/a$.

Thus a and $(G\bar{\nu})Q_t/b$ could be found individually. Since $G\bar{\nu}$ is obtained from measurements of M_0 as a function of modulation frequency $n/2\pi$, it follows that we obtain, from (A 6), a and Q_t/b , and the absorption coefficient $Cs\bar{\nu}$ from a . If, however, proportionality between M_0 and P persists at very large values of P , then it is possible to set lower limits only to $Cs\bar{\nu}$, the total absorption in nepers along the path s .

Appendix

An important quantity that appears in the theory of wave interaction as outlined in § 3 is W , the mean power given to a free electron in a gas by a high frequency alternating electric field $E_1 \cos p_1 t$, which for simplicity was written in equation (5) as $W = AE_1^2$. Expressions for W were derived by Bailey (1937), and independently at about the same time, slightly more accurate expressions were found by one of us (Huxley 1937). (Note that Bailey's w is $1/\nu$ times the W used in § 3. It is the work communicated per collision.) It is found that the high frequency conductivity of a gas containing N electrons per cm^3 is

$$\sigma_{\parallel} = \frac{2Ne^2}{3m} \left(\frac{\nu}{p_1^2 + \nu^2} \right) \left(1 + \frac{p_1^2}{p_1^2 + \nu^2} \right) \text{ E.M.U.} \quad \dots\dots (A 7)$$

$$\text{and} \quad \sigma_{\perp} = \frac{1Ne^2}{3m} \nu \left\{ \frac{1}{\nu^2 + (p_1 - \omega)^2} \left[1 + \frac{(p_1 - \omega)^2}{\nu^2 + (p_1 - \omega)^2} \right] + \frac{1}{\nu^2 + (p_1 + \omega)^2} \left[1 + \frac{(p_1 + \omega)^2}{\nu^2 + (p_1 + \omega)^2} \right] \right\}, \quad \dots\dots (A 8)$$

in which $\sigma_{\parallel}$ and $\sigma_{\perp}$ are the real parts of the conductivities parallel and perpendicular to the earth's magnetic field H and $\omega = H\mathbf{e}/m$ (E.M.U.).

The work done by the components of E_1 parallel and perpendicular to H is $W_{\parallel} = (\frac{1}{2}\sigma_{\parallel}E_1^2 \cos^2 \theta)/N$ and $W_{\perp} = (\frac{1}{2}\sigma_{\perp}E_1^2 \sin^2 \theta)/N$, where θ is the angle between E_1 and H . When $p_1^2 \ll \nu^2 \ll \omega^2$ as is the case in these tests, then

$$\sigma_{\perp} \rightarrow 0; \quad \sigma_{\parallel} \rightarrow \frac{2Ne^2}{3m} \left(\frac{\nu}{p_1^2 + \nu^2} \right) \left(1 + \frac{p_1^2}{p_1^2 + \nu^2} \right), \quad \dots\dots (A 9)$$

$$\text{and} \quad W = AE_1^2, \quad \dots\dots (A 10)$$

$$\text{with} \quad A = \frac{\cos^2 \theta}{3} \cdot \frac{e^2}{m} \left(\frac{\nu}{p_1^2 + \nu^2} \right) \left(1 + \frac{p_1^2}{p_1^2 + \nu^2} \right). \quad \dots\dots (A 11)$$

The following simplifications of (A 11) are applicable with the restrictions specified:

- (a) $p_1^2 \gg \nu^2$: $A \simeq 2e^2\nu \cos^2 \theta / 3mp_1^2$;
- (b) $p_1^2 \sim \nu^2$: $A \simeq e^2\nu \cos^2 \theta / 2m(\nu^2 + p_1^2)$;
- (c) $p_1^2 \ll \nu^2$: $A \simeq e^2 \cos^2 \theta / 3m\nu$.

In these special cases W is easily calculated from (A 10); the appropriate value is then used in (22).

REFERENCES

- BAILEY, V. A., 1937, *Phil. Mag.*, **23**, 744; 1938, *Ibid.*, **26**, 425.
 BAILEY, V. A., and MARTYN, D. F., 1934, *Nature, Lond.*, **133**, 218; *Phil. Mag.*, **18**, 369.
 GROSSKOPF, J., 1938, *Hochfrequenztechn. u. Elektroakust.*, **51**, 18.
 HEALEY, R. H., and REED, J. W., 1941, *Behaviour of Slow Electrons in Gases* (Amalgamated Wireless—Australasia), § 6.5.
 HUXLEY, L. G. H., 1937, *Phil. Mag.*, **23**, 210, 442.
 HUXLEY, L. G. H., FOSTER, H. G., and NEWTON, C. C., 1947, *Nature, Lond.*, **159**, 300.
 TELLEGEN, B. D. H., 1933, *Nature, Lond.*, **131**, 840.

The Diffraction Structure of the Elementary Coma Image

By R. KINGSLAKE

Eastman Kodak Company, Rochester, N.Y. *

MS. received 25 October 1947

ABSTRACT. The integrations reported in this paper demonstrate conclusively that the elementary coma image is of the size and shape determined by ordinary geometrical optics, and that the only effect of diffraction is to break up the image into an elaborate fine structure of dots and lines of light. Practically all the light in the image is confined within the triangular space between the principal ray at the tip of the figure and the sagittal focus. Between the sagittal and tangential foci there is a small amount of light broken up into a series of approximately concentric curved bands of darkness and light, centred about the brightest part of the image. Even when the amount of coma is very small, comparable to the Rayleigh Limit, the characteristic shape of a typical coma-image is already making its appearance. The accuracy of the theoretical predictions is fully and completely borne out by the actual photographs of a comatic image.

§ 1. INTRODUCTION

By the aberration "coma" is meant that third-order or Seidel aberration which leads to a wavefront differing from the ideal sphere by an amount P , given by

$$P = -ay(y^2 + z^2) \quad \dots\dots(1)$$

where y, z represent the coordinates of a point in the exit pupil of the lens, and a the amount of coma present; a varies linearly with the image height above the lens axis. The centre of curvature of the ideal or reference sphere is assumed to be the ordinary Gaussian extra-axial image point. In the absence of distortion, this is the point where the principal ray of the oblique pencil pierces the paraxial focal plane. The minus sign in (1) indicates that with positive coma, the wave is lagging for points above the middle of the lens.

Since we are discussing one of the Seidel aberrations, we must assume that both the aperture and the obliquity are very small. The longitudinal departure, P , of the wavefront in the exit pupil, relative to the reference sphere, amounts at most to a few wavelengths and is thus infinitesimal compared with y and z .

§ 2. GEOMETRICAL IMAGE

The y and z components of the point where a given ray pierces the paraxial focal plane, relative to the Gaussian image-point, are given by :

$$y' = -l \frac{\partial P}{\partial y}; \quad z' = -l \frac{\partial P}{\partial z} \quad \dots\dots(2)$$

where l is the longitudinal distance from the exit pupil to the image plane.

Thus

$$y' = al(3y^2 + z^2); \quad z' = 2alyz. \quad \dots\dots(3)$$

* The major part of the computations involved in this paper were made by the author in 1936-37 when he was on the faculty of the Institute of Applied Optics, University of Rochester.

These quantities are best interpreted by expressing the aperture-point position in polar coordinates. Writing

$$y = r \cos \theta; \quad z = r \sin \theta \quad \dots\dots(4)$$

the components of the ray-point displacement become

$$y' = alr^2(2 + \cos 2\theta); \quad z' = alr^2(\sin 2\theta). \quad \dots\dots(5)$$

The values of y' and z' , for a single zone of the aperture, give a double circle of radius alr^2 in the image plane as shown in figure 1, the centre of the circle

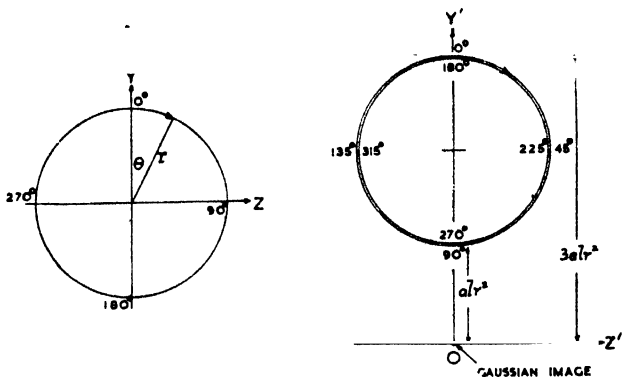


Figure 1. Ray-point in exit pupil and in image plane.

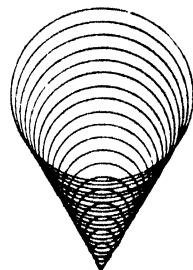


Figure 2. The geometrical coma image.

being at a height $2alr^2$ above the Gaussian image-point. Each elementary zone of the lens forms such a double circle, and the complete family of double circles for the circular lens aperture fit between two common envelope lines, at 60° to each other, which intersect at the Gaussian image or principal-ray point (figure 2).

The top and bottom ("tangential") rays of any zone intersect at the top-most point of the double circle in figure 1, and the front and rear ("sagittal") rays of a zone meet at the lowest point of the double circle. If the height of the tangential image above the principal ray is coma_T , and the height of the sagittal image coma_S , then

$$\text{coma}_T = 3 \text{coma}_S. \quad \dots\dots(6)$$

§ 3. THE PHASE RELATIONSHIPS OF COMA

The phase difference between light arriving at the Gaussian image-point from the point (y, z) and from the centre of the exit pupil is given by equation (1), since phase $= 2\pi P/\lambda$. If, however,

we shift our reference-point to some point vertically above the Gaussian image, we add a further phase difference which can be used to simplify computation of the light distribution. From figure 3 it is seen that the phase difference at D between light travelling along AD and along PD is given by :

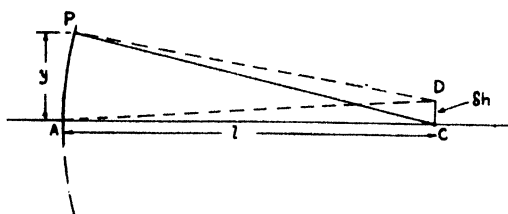


Figure 3. The path difference introduced by a small vertical shift of the reference point.

$$\frac{2\pi}{\lambda} [PD - AD] \simeq \frac{2\pi}{\lambda} \left[\frac{y}{l} \cdot \delta h \right] \text{ radians.} \quad \dots\dots(7)$$

Hence from (1) and (7), the phase-difference between light arriving at *any* reference-point on the centre-line of the image from (y, z) and from the middle of the aperture is $(2\pi/\lambda)[-ay(y^2 + z^2) + (y/l)\delta h]$.

Since for any single zone in the exit-pupil aperture x is negligible compared with y or z , this may be written

$$(2\pi/\lambda)(y/l)[-alr^2 + \delta h]. \quad \dots\dots(8)$$

The phase-difference will be zero for *one entire zone* if

$$\delta h = alr^2 \quad \dots\dots(9)$$

and, since alr^2 is equal to the sagittal coma of the zone, we reach the very important conclusion that *each separate zone of the lens forms a perfect equiphase focus at the sagittal image of the zone*. Hence to determine the light distribution in the focal plane, for a complete lens, we divide the lens aperture into a number of circular zones of equal area, and integrate the amplitude contributions of all the zones at each of a number of discrete points in the focal plane. Contours of light intensity can then be drawn through the various points so computed.

§ 4. THE AMPLITUDE DISTRIBUTION IN THE IMAGE OF A POINT FORMED BY A SINGLE NARROW CIRCULAR ZONE OF THE LENS

This problem is classical, and a good treatment has been given by Martin (1930), who shows that the amplitude at a radial distance t from the centre of the image of a single zone of radius r is given by

$$J_0\left(\frac{2\pi}{\lambda} \cdot \frac{tr}{l}\right). \quad \dots\dots(10)$$

For convenience in computation and in plotting the results, we shall assume that the radius of the lens aperture is unity. We can then introduce a symbol $W = 2\pi t/\lambda l$ to represent the radial distance of our integration point R from the centre of a zonal image. By comparing (10) with (7) it may be seen that the physical meaning of W is the phase difference between light reaching the point R from the edge and centre of the (unit) marginal zone of the lens aperture. In the case of coma, there is no path difference for any zone at its sagittal focus, and here W is zero. Thus the amplitude contribution of a zone of radius r , at a radial distance W (radians) from the centre of the zonal image, is given by

$$J_0(Wr). \quad \dots\dots(11)$$

§ 5. RADIAN MEASURE OF THE AMOUNT OF COMA IN THE LENS

The amount of coma present in the lens may be expressed by a type of radian measure similar to that used in the W notation introduced in § 4 by writing

$$B = (2\pi/\lambda l)(al) = 2\pi a/\lambda. \quad \dots\dots(12)$$

By comparing (1) with (12), it is clear that B represents the phase difference, in radians, between light reaching the Gaussian image-point from the top and centre of the lens aperture.

These two quantities, B and W , may therefore be used to replace coma_s and t respectively, with complete assurance that the units will be properly compatible.

§ 6. THE RAYLEIGH LIMIT FOR COMA

Conrady (1929) has stated that the Rayleigh Limit for Seidel coma is reached when

$$\text{coma}_s = 0.5 \lambda / \sin U'_M \quad \dots\dots (13)$$

which becomes $\lambda/2$ for a lens of unit radius. Converting this into the B notation, we find that :

$$B\text{-tolerance} = (2\pi/\lambda) \cdot (\lambda/2) = \pi \text{ radians.}$$

Now Rayleigh based his limit on the fundamental idea that the aberration in a lens will be acceptable if it is just sufficient to cause a diminution of the peak intensity in the image to 80% of its value for a perfect lens. It is therefore of interest to plot the variation of intensity along the centre-line of the image when $B = \pi$ to see how closely Conrady's statement is borne out.

The amplitude at a point on the centre-line of the image, distant Q (radian measure) from the Gaussian image-point, can be found by integrating the zonal contributions given by equation (11). The integration must be taken over a set of infinitesimal zones of equal area, i.e. of equal increment in r^2 . In figure 4,

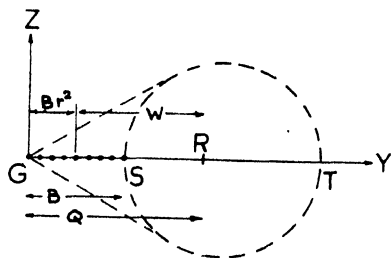


Figure 4. Procedure for integrating the light distribution along the centre-line of the coma figure.

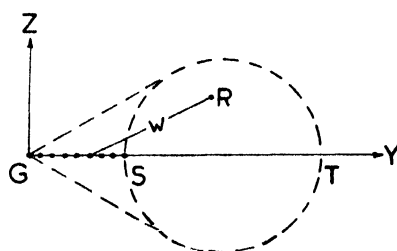


Figure 7. The distance W' from a zonal focus to the point of integration, R .

the heavy dots represent the sagittal foci of such a series of zones, equally spaced along the mid-line since the r^2 of the zones are equally spaced; G is the Gaussian image-point, S the sagittal focus of the marginal zone for which $r=1$, and R the point at which the net amplitude is to be determined by integration. Since the distance GS represents B , and GR represents Q , the distance of R from the equiphase focus of any zone r is $W = (Q - Br^2)$ radians. The total amplitude is thus, by equation (11),

$$\int_0^1 J_0(Wr) d(r^2) = \int_0^1 J_0[r(Q - Br^2)] d(r^2). \quad \dots\dots (14)$$

This equation has been given by Steward (1926 a, b) and Buxton (1926).

A series of values of (14) were calculated for a succession of Q values, for $B = \pi$ and by plotting graphs of the J_0 values against r^2 and integrating with a planimeter, the central section of the coma image could be plotted (figure 5). It is clear that unit intensity can be obtained only when $B=0$ and $Q=0$, that

is, at the centre of a perfect Airy disc. It will be seen from figure 5 that Conrady's tolerance is not in accordance with Rayleigh's rule, for the maximum intensity is about 89% of the central Airy disc intensity when $B = \pi$.

Steward's (1926 a) data for a central section of intensity of the coma image for the case $B = 4$ are shown in figure 6. This case does actually represent the Rayleigh limit, for the peak intensity is now 80%. Conrady's formula (13) above should therefore be emended to read

$$\text{Rayleigh limit for coma}_s = 0.64 \lambda / \sin U'_M.$$

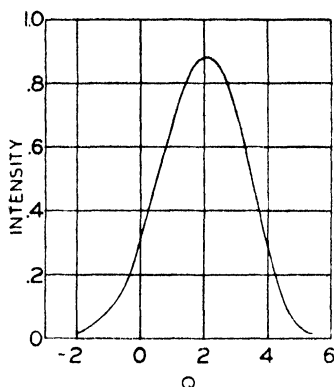


Figure 5. The centre-line section of the coma image, for $B = \pi$.

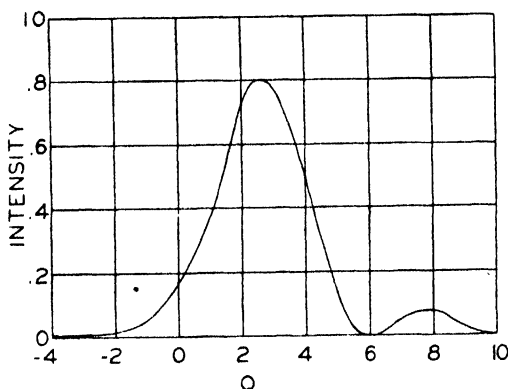


Figure 6. The centre-line section of the coma image, for $B = 4$ (from Steward).

§ 7. LIGHT CONTOURS

Light contours for the case $B = 4$

To determine the light distribution over the whole image, we assume that the lens aperture has unit radius, and divide the aperture into a number of zones equispaced in r^2 . The procedure follows that suggested by Conrady (1919), except that he used Simpson's rule for the mechanical integration instead of a planimeter. A point R is taken in the focal plane (figure 7), and the oblique distances W from R to each of the equiphase foci of the various zones are tabulated. $J_0(Wr)$ is then plotted against r^2 and the area of the graph determined in the usual way with a planimeter. This area is the net amplitude at the point R due to the superposition of the contributions of all the zones in the entire lens. By taking a succession of points R in the image plane, a number of longitudinal and cross sections of the light distribution (in amplitude) can be plotted. The points where these sections cross the desired contour lines of intensity can then be joined with smooth curves.

For the case of $B = 4$, the lens was divided into eight zones having radii r equal to $\sqrt{(1/8)}$, $\sqrt{(2/8)}$, $\sqrt{(3/8)}$, . . . 1.0 respectively. The amplitude-contribution graphs were plotted for each of about 150 points in the image plane, having Y values lying between -10 and $+14$ (radian measure) and Z values lying between 0 and 8. As the pattern is symmetrical about the Y axis, only one side need be computed. Each curve was then integrated with a planimeter giving directly the net amplitude at that particular point. The results of the integrations are summarized in table 1.

Table 1. Computed amplitudes for $B=4$

$Z \backslash Y$	0	± 1	± 2	± 3	± 4	± 5	± 6	± 8
-10	0							
-9	-0.008							
-8	-0.002		-0.006					
-7	0.002							
-6	0.020		0.024		-0.004	-0.002	-0.012	
-5	0.044			0.038	0.028	0.014	0.0	
-4	0.064		0.060	0.070	0.048	0.040	0.016	
-3	0.078	0.078	0.070	0.052	0.048	0.052	0.036	
-2	0.104	0.088	0.066	0.038	0.036	0.040	0.052	
-1	0.218	0.186	0.124	0.052	0.004	0.008	0.034	
0	0.426	0.368	0.218	0.078	-0.016	-0.046	0.014	
1	0.648	0.580	0.376	0.134	-0.028	-0.086	-0.032	
2	0.848	0.734	0.474	0.168	-0.048	-0.118	-0.070	0.060
3	0.888	0.772	0.472	0.158	-0.078	-0.156	-0.110	0.064
4	0.710	0.584	0.342	0.052	-0.142	-0.186	-0.110	0.066
5	0.366	0.294	0.106	-0.106	-0.212	-0.190	-0.090	0.090
6	0.040	-0.018	-0.128	-0.218	-0.236	-0.158	-0.028	0.100
7	-0.208	-0.230	-0.246	-0.246	-0.194	-0.078	0.032	0.110
8	-0.264	-0.258	-0.240	-0.176	-0.088	0.024	0.102	0.088
9	-0.152	-0.148	-0.096	-0.026	0.046	0.106	0.136	0.046
10	0.020	0.012	0.054	0.106	0.130	0.144	0.130	-0.008
11	0.134	0.118	0.152	0.160	0.136	0.100	0.058	-0.052
12	0.130	0.128	0.128	0.116	0.070	0.038	-0.018	
13	0.076	0.050	0.052	0.020	-0.020			
14	-0.034	-0.044		-0.052				

The centre-line amplitudes are a recomputation of those given by Steward and shown in figure 6.

A series of longitudinal and transverse sections were then plotted from the above figures, and contour points were taken from the section graphs at amplitude values $\sqrt{I}$ corresponding to intensities $I=0, 2, 4, 6, 8, 10, 20, 30, 40, 50, 60, 70, 80\%$.

The resulting contour map of the light distribution is seen in figure 8.* In this figure, the zero-intensity loci are shown dotted, and the outline of the geometrical image is included.

Light contours for the case $B=3$

It has recently been brought to my attention that the intensity contours for the case of $B=3$ have been very fully worked out by Nijboer (1946). This plot is reproduced in figure 9 as it may not be readily available.

Light contours for the case $B=6.14$ radians (352°)

This case has been integrated by Martin (1922 a), and the central part of the pattern plotted. It was later verified photographically by the use of a specially designed microscope objective having the desired amount of coma (1922 b).

Light contours for the case $B=20$

For this case, 16 zones were taken of radii $\sqrt{(1/16)}, \sqrt{(2/16)}, \dots, 1$, respectively, the equiphase-points of the successive zones being, therefore, 1.25 (radian) units apart. The amplitude-contribution graphs were plotted for some 190 points in the focal plane, and each graph was integrated as before. The resulting net amplitudes are tabulated in table 2. The contour pattern (figure 10) was plotted as before by drawing a series of longitudinal and transverse sections and marking the points for the amplitudes $\sqrt{I}$ corresponding to $I=0.5, 1, 2, 4, 6, 8, 10\%$.

* This pattern differs somewhat from that given by Steward (1926 a, p. 156); there are also small differences between the amplitude distribution along the centre-line of the pattern as given in table 1 and as given by Steward.

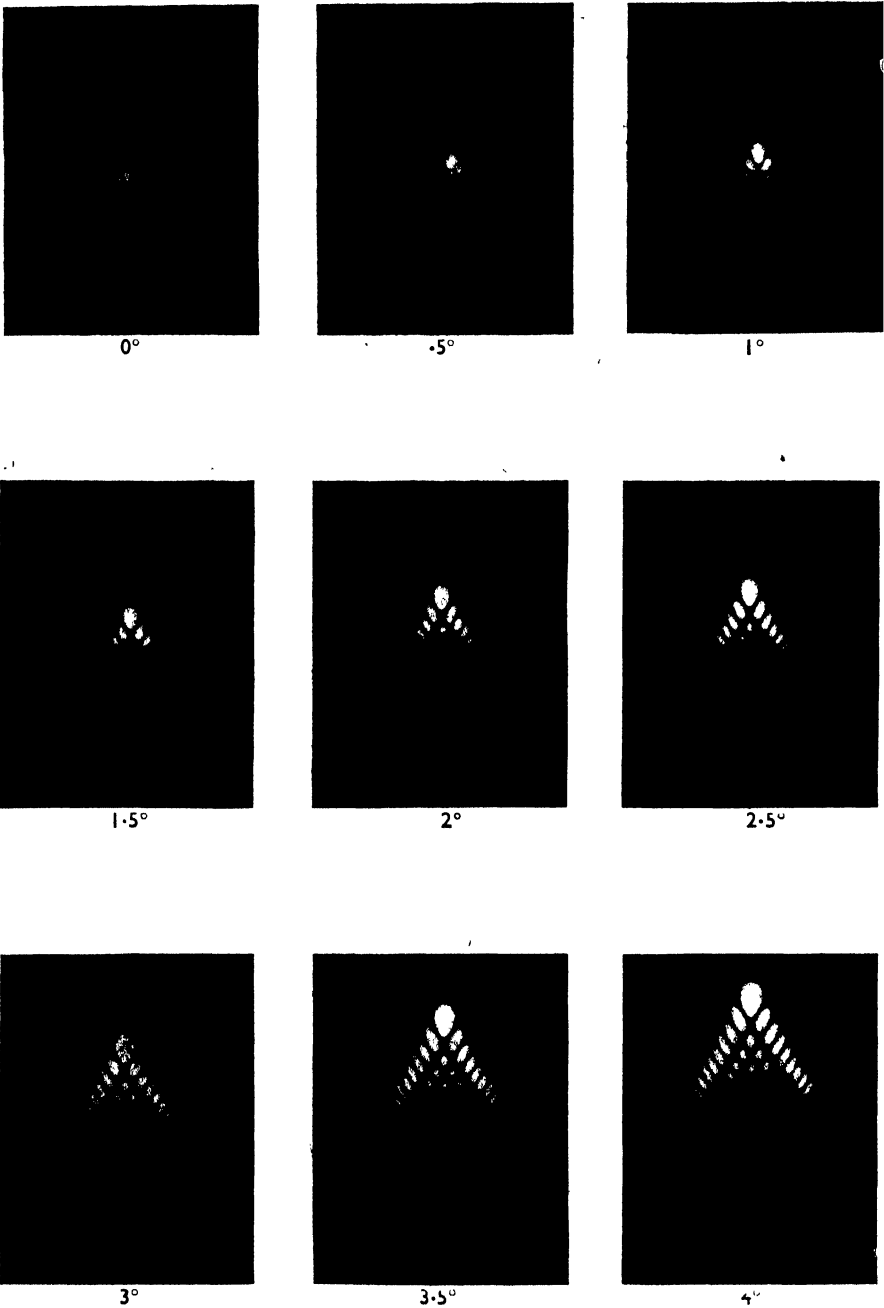


Figure 13. Photographs of a comatic star image showing progressively increasing amounts of coma.

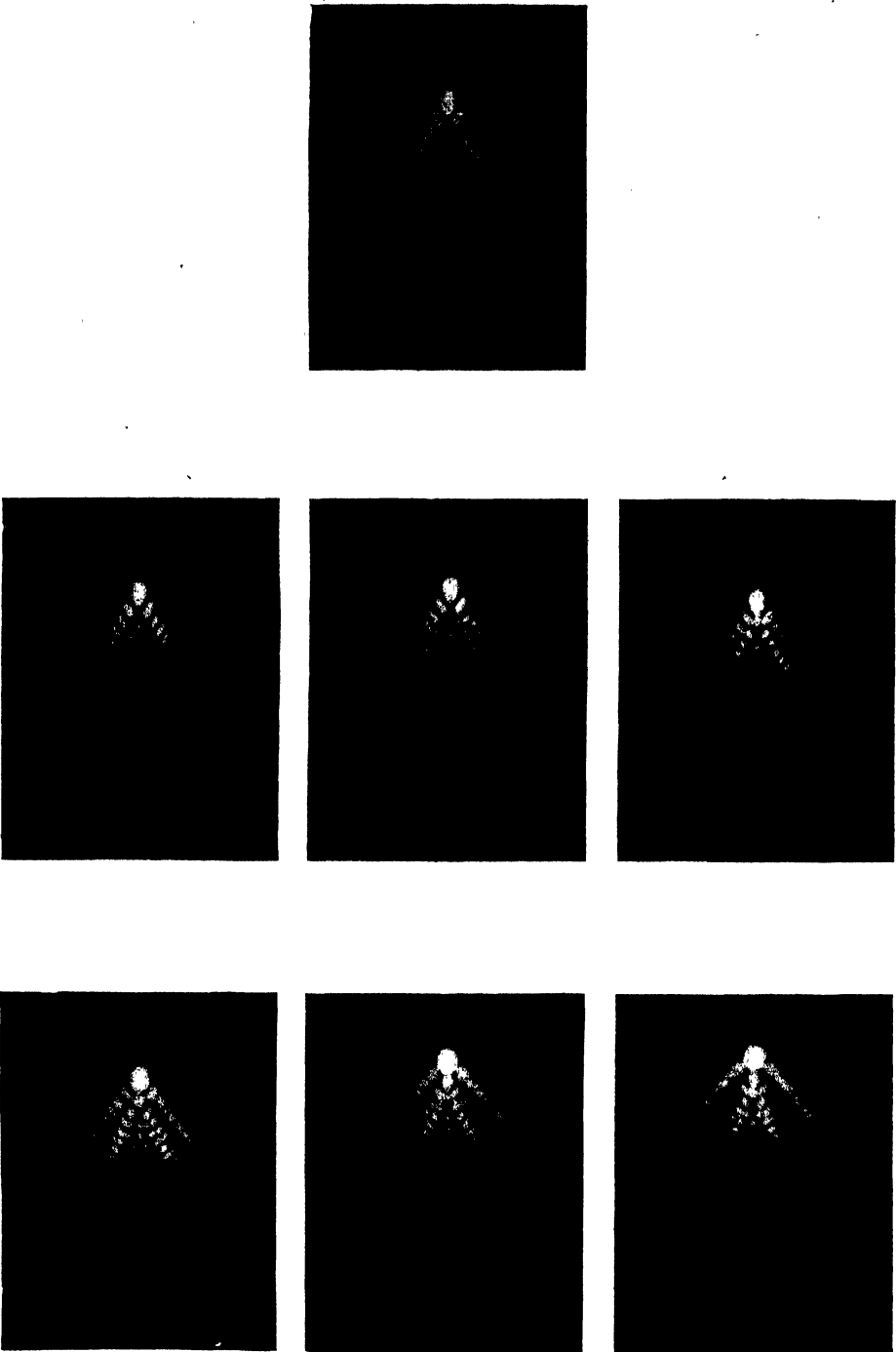


Figure 14. Photographs of a comatic star image showing the effects of progressively increasing departures from the focal plane.

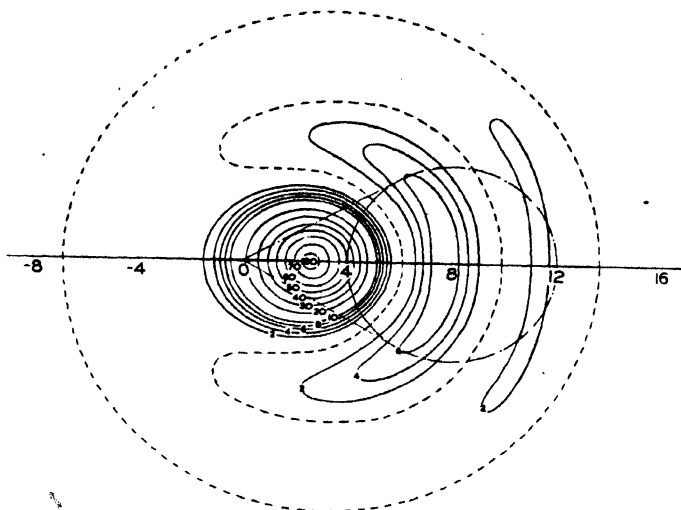


Figure 8. Contours of intensity for $B=4$

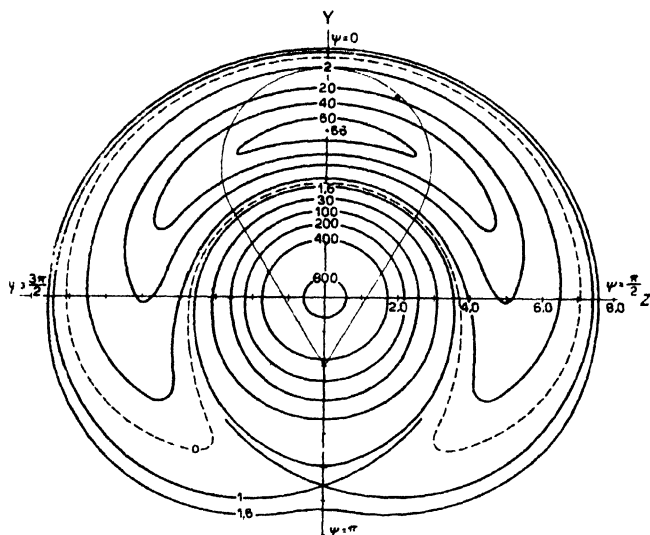


Figure 9. Contours of intensity for $B=3$ (from Nijboer).
Reproduced by kind permission of the Elsevier Publishing Co., Inc.

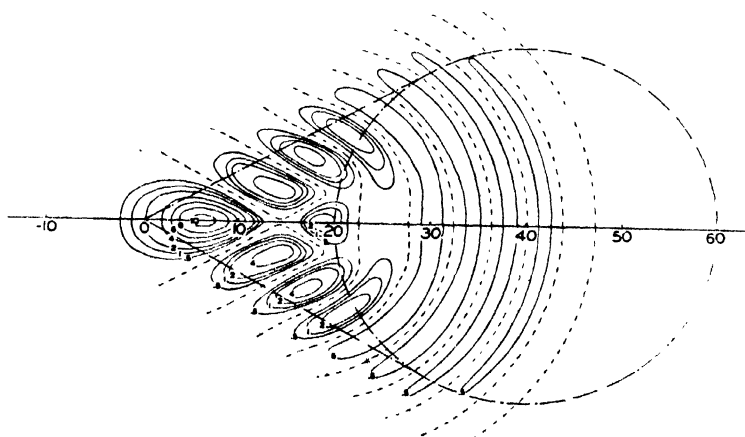


Figure 10. Contours of intensity for $B=20$.

As in the previous cases, the loci of zero intensity are shown dotted, and the outline of the geometric coma pattern is included for reference. The pattern was computed to $Y = 50$, which is just beyond the centre of the circle in the geometrical image belonging to the marginal zone.

Light contours for the case $B = 40$

For this case, the number of zones was increased to 32, of radii $\sqrt{(1/32)}$, $\sqrt{(2/32)}$ etc. the equiphase-points of successive zones being 1.25 (radians) apart as for the $B = 20$ case. This was a convenience in computation since the oblique distances Q from all the equiphase foci to the integration points had already been evaluated. The labour of computing the net amplitude at each integration point was considerable, as it involved computation of 32 four-digit products of r and Q , determination of the J_0 curve for each product and integration by planimeter. To show how rapidly the amplitude contributions alternate from one zone to another, a few examples are included in figure 11.

In all, the net amplitudes at some 256 points in the image plane were computed in this way (table 3). The longitudinal and cross section graphs were then plotted, and contours were drawn (figure 12) at amplitude levels corresponding to intensities 0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4%.

§ 8. PHOTOGRAPHIC CONFIRMATION OF THE COMPUTED PATTERNS

The problem here is to obtain a lens in which there is a considerable amount of coma and negligible spherical aberration and astigmatism. Such a lens might be, for example, the rear half of a symmetrical system designed to give excellent anastigmatic definition at unit magnification, with parallel light entering through the central stop.

Fortunately a suitable lens of this type was available. The rear half had a focal length of about 8 inches and an aperture of $f/5.6$, the field out to 5° from the axis being flat and highly corrected for astigmatism. The spherical aberration was so small that even at $f/5.6$ the axial image in monochromatic light was a perfect Airy-disc.

With this lens set at 4° obliquity, an attempt was made by adjusting the iris diaphragm to reproduce as closely as possible the computed pattern for the

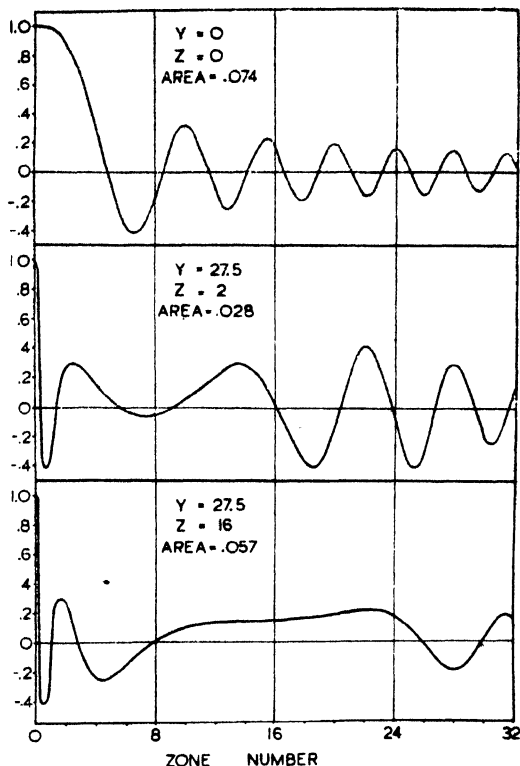


Figure 11. Some typical amplitude-contribution curves for the case $B = 40$.

Table 3. Computed amplitudes for $B=40$ [illegible]

case $B=40$. Then, since coma is proportional to image height, it was expected that each degree of obliquity would represent 10 units of B . However, this operation proved to be surprisingly difficult, and in the actual series of enlarged photographs of the star image at every half degree of obliquity, reproduced in figure 13, it will be seen that the computed patterns for $B=4, 20$, and 40 very closely resemble the photographs at $\frac{1}{2}^\circ$, 2° , and $3\frac{1}{2}^\circ$ respectively. The resemblance between the computations and the photographs is remarkably close, even to the shape of the individual little spots and bands of light in the images.

The original star-image photographs were made through an 8 mm. apochromatic microscope objective on ordinary 16 mm. motion-picture panchromatic film, in sodium D light, the negatives being then further enlarged in making the prints.

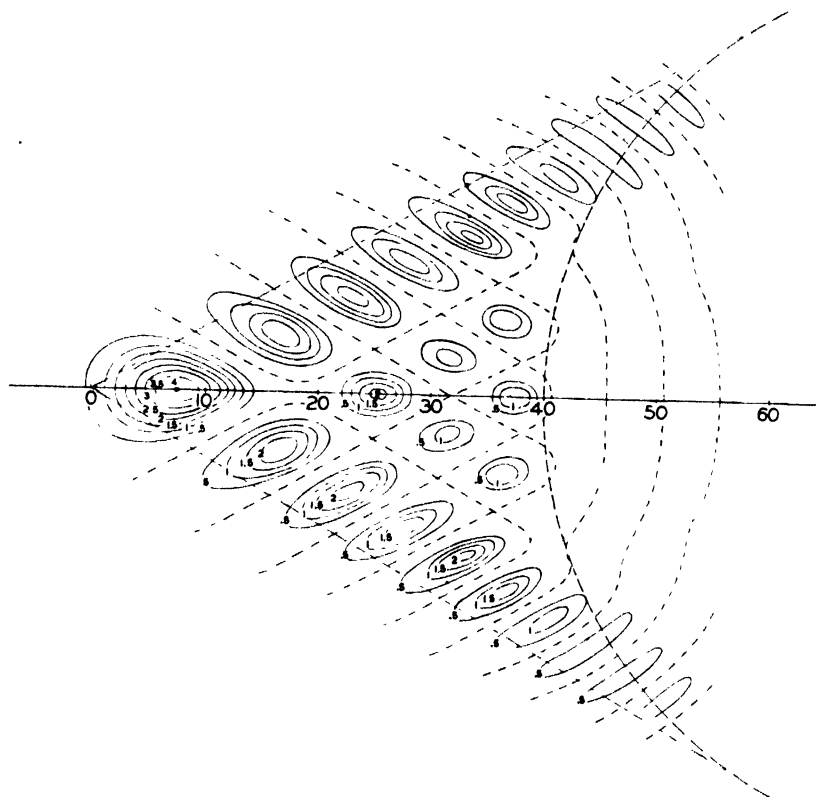


Figure 12. Contours of intensity for $B=40$.

§ 9. OUT-OF-FOCUS COMA IMAGES

The problem of computing the light distribution in an out-of-focus coma image is very difficult, as the convenient fact that each zone forms a sharp equiphase focus at its sagittal image can no longer be utilized. Consequently, use was made of the lens system mentioned in the last paragraph to provide this information. The lens was tilted to $4\frac{1}{2}^\circ$ obliquity (corresponding approximately to $B=50$), and a succession of star-image photographs were made for a series of focal settings at successive intervals of 0.1 mm. out of focus. It was found to be immaterial whether the image plane was chosen within or beyond the best

focus plane, thus again confirming the very high degree of correction of all the aberrations other than coma. These out-of-focus star images are shown in figure 14.

REFERENCES

- BUXTON, A., 1926, *Proc. Opt. Conv.*, II, 771.
 CONRADY, A. E., 1919, *Mon. Not. R. Astr. Soc.*, **79**, 577 ; 1929, *Applied Optics and Optical Design*, I, 395.
 MARTIN, L. C., 1922 a, *Mon. Not. R. Astr. Soc.*, **82**, 310 ; 1922 b, *Trans. Opt. Soc.*, **24**, 1 ; 1930, *Applied Optics*, I, 95.
 NIJBOER, B. R. A., 1946, Part of thesis embodying the coma analysis—see Bouwers, A., *Achievements in Optics*, Monographs on the Progress of Research in Holland (New York and Amsterdam : Elsevier Publ. Co., Inc.).
 STEWARD, G. C., 1926 a, *Phil. Trans. Roy Soc. A*, **225**, 153 ; 1926 b, *Proc. Opt. Conv.*, II, 791.

Diffraction and Optical Image Formation

By F. ZERNIKE

University of Groningen

The Thomas Young Oration, delivered 24 September 1947

IF one reviews the historical development of this subject, one is struck by the very slow development of the fundamental concepts and methods. The wave theory of light seemed not only well established, but even nearly completed, by about 1820, after the work of Young and Fresnel. However, the first application to the resolving power of optical instruments was made by Airy in 1835, the second by Helmholtz and by Abbe forty years later. The next step, the extension to the case of lens errors, was not made until about 1900 by Strehl, and 1920 by Conrady and by Richter. A somewhat different branch, that of coherence, begun hesitatingly by Verdet in 1860 and developed by Michelson in 1890, found its practical application about 1920 at Mt. Wilson, and its further theoretical foundation by van Cittert (1934). You will understand that I have often been astonished to find that problems of such old standing still showed themselves open to further development.

I shall treat of three different, though interconnected, subjects: the coherent background, the degree of coherence and the diffraction theory of aberrations. As an introduction, I would demonstrate a few diffraction experiments. By the aid of a small arc lamp, condenser and vertical slit I throw the shadow of a thin vertical needle on the screen. In order to make the details visible through the whole room, I enlarge the shadow, in the horizontal direction only, by a short focus *cylindrical* lens. The external fringes are clearly seen, but also the internal fringes, which Thomas Young explained by interference between the two beams diffracted at the edges (figure 1 (a)). With a wedge-shaped needle (figure 1 (b)) the different behaviour of the two kinds with increasing thickness of the needle is seen at a glance. The internal fringes soon become too dark to be seen. Yet with another thicker needle they reappear very clearly, a dark fringe at the centre (figure 1 (c)). I have used a simple trick here: the needle is double, through

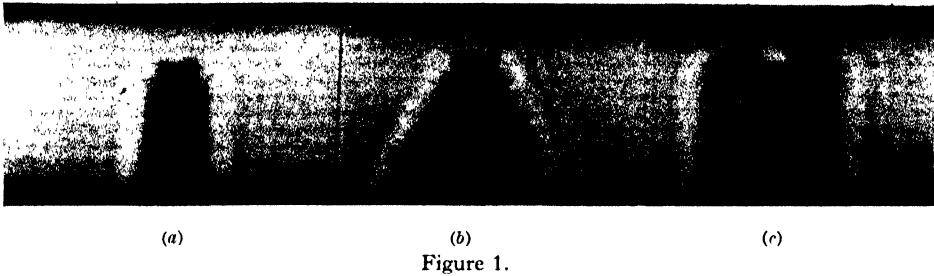


Figure 3. Diffraction figure of a slit, *above* without background, *middle* with coherent background, *below* background alone.



Figure 4. F. G. Pease at the eyepiece of his 15-metre stellar interferometer

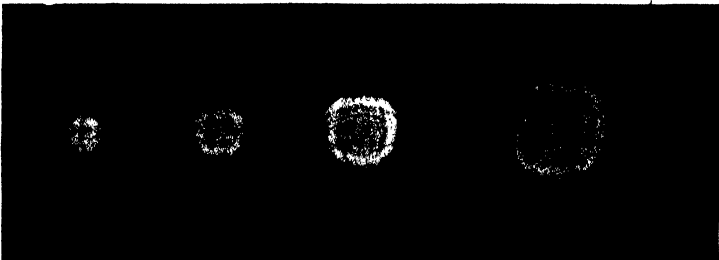


Figure 5. Astigmatism patterns for $\beta=8, 15, 26$ and 45 respectively.

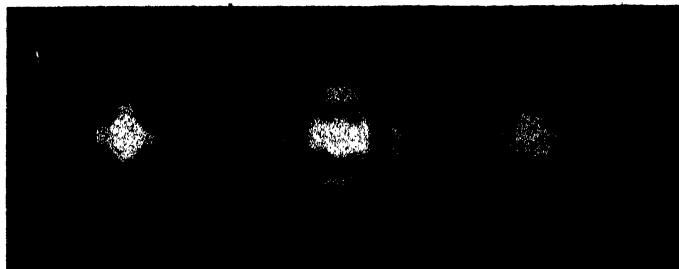


Figure 6. Astigmatism for $\beta=1$, *left* without background, *middle* same with background, *right* background alone.

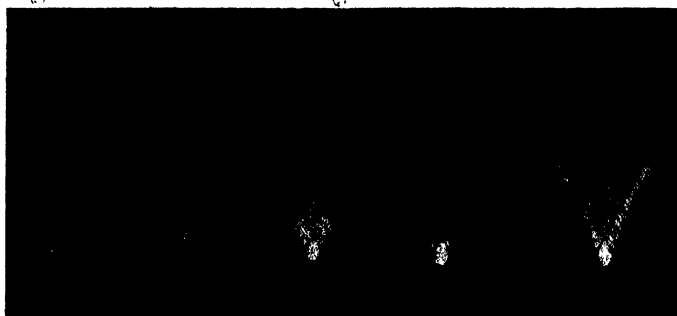


Figure 7. Coma patterns with $\beta=0.7, 2.5, 10, 20$ and 50 respectively.

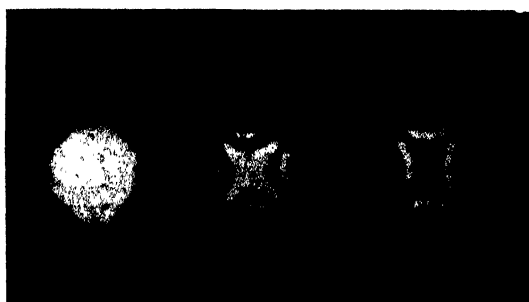


Figure 8. Astigmatism with $\beta=12$, without background, with background in phase and with background one quarter behind.

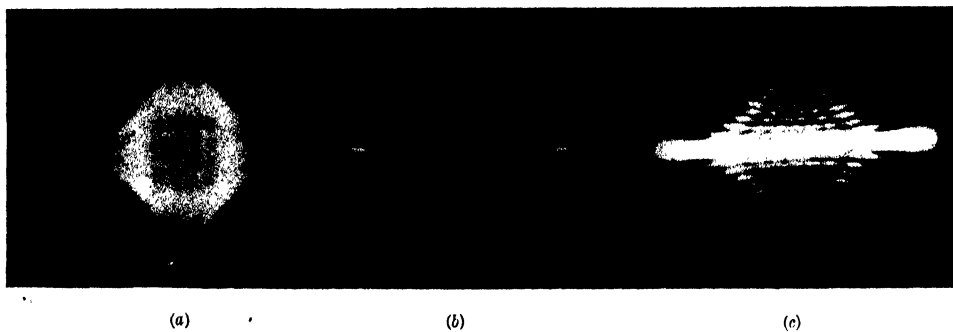


Figure 9. Astigmatism, $\beta=17$, (a) complete pattern, (b) edge effect alone, (c) pattern at focal line.

the narrow slit between some light passes and by diffraction is spread over the dark shadow. Evidently, the background thus created is *coherent* with the faint light of the fringes. Therefore the latter are effective here with their *amplitude*. A simple calculation shows that the fringes are much more easily seen in this way.

In order to obtain a similar effect with the light bent into the shadow at a single edge, I use a different artifice: the shadow-throwing screen is made slightly transparent. A number of fringes is now seen in the shadow (figure 2). In general it is found advantageous to combine both methods. Thus the Fraunhofer diffraction image of a slit was thrown on a coherent background in the following way: The slit is covered by a plane-parallel glass covered with a transparent layer of a strongly absorbing metal (silver or aluminium). In this layer a narrow scratch is made exactly in the centre of the slit. The extra light through this scratch gives rise to a much broader, but not too much weaker, diffraction image, the central fringe of which may cover the whole diffraction image of the slit (figure 3).

In these cases one can in different ways also change the phase of the background at will. The result is therefore that the amplitudes and phases at various points of a diffraction image can be observed experimentally by aid of the coherent background.

In this method there is further a clear parallelism with theory. In order to define clearly the meaning of the phase difference between points of a diffraction image, the theorist must be sure to introduce a *surface of reference*, which may be plane, cylindrical, spherical, etc. in a more or less arbitrary way. Indeed, controversies have sometimes arisen through neglect of this. In the same way the experimenter must introduce an *auxiliary coherent wave*, which may radiate from a more or less arbitrary point or line.

The question of *partial coherence* started with Verdet, who asked at what distance apart two points on a screen illuminated by the sun would still have coherent vibrations. He found that this was determined by the apparent diameter of the sun, the actual distance being less than 1/20 millimetre. Let us imagine an experimenter who wants to verify this. He will take a piece of tinfoil and prick very small holes in it, in pairs of various distances. Through each hole a cone of light will pass and adjacent cones will overlap and show interference fringes. But if they do not, shall we call the adjacent cones *incoherent*? I agree, but I must warn you that this is quite a daring step, undertaken only recently. Indeed, opticians have been very cautious, calling vibrations incoherent only when they came from different sources, thus making sure that their haphazardly changing phases would be statistically independent. Well, as soon as we take the more daring point of view, a new opportunity presents itself, namely to call vibrations partially coherent when they give fringes of lower visibility and to define their *degree of coherence* γ to be equal to the number between 0 and 1 which expresses their visibility:

$$\text{visibility} = \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}} = \text{degree of coherence } \gamma.$$

This new concept of degree of coherence leads to various remarkable results. For our sunlit screen γ vanishes at a distance of 0.07 mm., but at larger distances it rises again and goes up and down many times. More exactly, its course is

equal to that of the amplitude of the diffraction image of a star formed by a telescope objective of angular aperture equal to the sun's diameter (Airy disc with rings). It would carry us too far from our subject to discuss the application of the same theorem (discovered by van Cittert 1934) to the rôle of a microscope condenser or of a condenser in front of a spectroscopy slit.

Another general theorem, on the contrary, is directly connected with image formation. It is the theorem about the propagation of γ , which states that a knowledge of the distribution of intensity, of degree of coherence and of relative phases in any surface intersecting a beam of light enables us to calculate the same quantities at a following or at a preceding surface. For instance, the image that will be formed in a photographic camera—i.e. the distribution of intensity on the sensitive layer—is present in an invisible, mysterious way in the aperture of the lens, where the intensity is equal at all points, namely in the distribution of γ in this aperture. And if you ask for the mathematical connection between the two: one is the Fourier transform of the other. This was probably known, in an incomplete form, to Michelson in 1890, but for lack of the requisite term and even of the requisite concept, he could not adequately express it. In such cases one feels the truth of E. Mach's statement that science serves to economize thinking. Indeed a single term may stand for a whole theory, may convey its ideas, theorems and concepts.

Michelson (1890) had suggested that it is possible to obtain the apparent diameters of stars from the visibilities of interference fringes, or, in our terminology, to find them from a determination of the degree of coherence as a function of distance apart. In the practical execution of this idea at the Mt. Wilson Observatory, our minute holes in the tinfoil were represented by two eight-inch mirrors under 45° which were movable along a six-metre steel beam, mounted across the opening of the 100-inch telescope. The mirrors reflect the light from a star towards the centre of the beam. Two fixed mirrors mounted there throw the light into the telescope. The observer first sees two star discs which he brings into coincidence by adjusting the mirrors. He then estimates the visibility of the fringes that appear and repeats this for various distances of the first mirrors. F. G. Pease, the designer of the 100-inch telescope, is to be credited for most of this work. It may be said that he devoted much of his time during the last fifteen years of his life to measuring degrees of coherence—without knowing it. All the same he measured various star-diameters and "resolved" Mizar, before only known as a spectroscopic double star (distance 0.011"). Figure 4 shows Dr. Pease at the eyepiece of a specially constructed 15-metre instrument. Perhaps future astronomers will build larger instruments of this kind, say of 50 or 100 metres, which will indirectly show details down to one thousandth of a second of arc.

This indirect method of studying celestial objects may well be compared with the study of crystal structure by x-ray diffraction. There also the synthesis from the Fourier transform to the image cannot be obtained by the direct optical method, that is, the image cannot be seen but must be calculated.

I shall dwell somewhat longer on my last subject, the rôle of diffraction in image formation in the presence of lens errors. The prevailing attitude among opticians was even recently the following. Diffraction causes a certain unavoidable deterioration of the ideal point image, which may be expressed by the radius a of the Airy disc. Any aberration present will also give a certain diffusion,

expressible by a "radius" b of the corresponding geometrical pattern. Both together may then be assumed to give a diffusion $a+b$. There is one thing in favour of this crude estimate: the tolerances for aberrations deduced from it are very much on the safe side.

My own work on the change of the Airy pattern caused by small aberrations was started in 1934 and completed by Nijboer in his thesis of 1942. A short survey of the results must suffice here. To begin with, the aberration will of course be expressed, not in terms of light rays and their intersection with the receiving plane, but as deviations of the wave surface in the exit pupil from the ideal spherical form. Let these deviations be represented by $V(y, z)$ as a function of the plane rectangular coordinates y, z in the circular opening, or as $V(r, \phi)$ in polar coordinates, such that $r=1$ at the edge. Secondly, this characteristic function V is developed in a series of *orthogonal polynomials*,

$$V(r, \phi) = \sum \beta_{nm} R_n^m(r) \cos m\phi, \quad \dots\dots(1)$$

which were specially constructed for the purpose. In the receiving plane with polar coordinates ρ, ψ the resulting diffraction image has then, to a first approximation, the amplitude

$$A(\rho, \psi) = 2\rho^{-1} \{J_1(\rho) - \sum i^{n+1} \beta_{nm} J_{n+1}(\rho) \cos m\psi\}. \quad \dots\dots(2)$$

Nijboer also gives the general form of the terms up to the fourth powers of the coefficients β .

At the centre of the pattern, the point of maximum amplitude for small errors, the result is

$$A_0 = 1 - \sum \beta_{nm}^2 / 4(n+1). \quad \dots\dots(3)$$

The intensity at the centre, A_0^2 , is a good measure of the quality of the image (Strehl's "definition-brightness"). A diminution of 10% may well be tolerated, $A_0 = 0.95$. Suppose, for instance that there is only ordinary spherical aberration, β_{40} must then be less than 1, whereas considered geometrically, this value would give a circle of least confusion of radius 6, or 3.1 times the radius of the Airy disc. Another remarkable advance lies in the fact that there are no mixed terms in (3), so that a higher aberration cannot be improved by small amounts of a lower one. In other words, the *balancing* of aberrations has been completely attained by our introduction of orthogonal polynomials.

The method described was thus fully successful for small errors. Dr. Nijboer found it increasingly difficult for larger aberrations. I can show a slide with the pattern for astigmatism with $\beta_{22} = 4$. The formula for this case fills a whole page. As it appeared hopeless to get any further theoretically, especially to get an insight into the gradual transition to the geometrical pattern for increasing errors, we turned to experiment. Let me give a few details of the way we obtained pure third order errors.

Astigmatism was obtained with a symmetrical biconvex lens of 1 m. focus and magnification one, thus excluding coma, turned into an oblique position through measurable angles up to 15° . The circular diaphragm remained fixed, perpendicular to the beam. The astigmatism is proportional to the square of the angle.

Coma was obtained from an ordinary achromatic telescope objective by shifting its components laterally in opposite directions. The coma is proportional to the shift and to the third power of the aperture.

Spherical aberration was obtained from a single meniscus lens in a reversed position, the amount being changed by changing the aperture. It is proportional to the fourth power of the aperture.

In all cases green mercury light was used and the diffraction pattern enlarged 5-10 times by an auxiliary lens. The largest amounts of aberration used could always be determined experimentally from geometrical optics (axial distance of focal lines etc.). Let me show a few examples of astigmatism and of coma (figures 5, 6 and 7). Mr. Nienhuis, who made these experiments and whose thesis is to appear in a few months, now proceeded to the experimental investigation of the formation of these patterns.

As an example I take the largest coma pattern. The wave retardation, expressed in radians, is in this case

$$V(r, \phi) = 3\beta r^3 \cos \phi = 3\beta y(y^2 + z^2), \quad \dots\dots(4)$$

there being no need here for the polynomial. Let us start from the geometric optical pattern. As is well known, each zone of the lens aperture gives rise to a circle with displaced centre and which is *described twice*. In fact, the optical path to the point (η, ζ) in the receiving plane becomes

$$V(y, z) - y\eta - z\zeta, \quad \dots\dots(5)$$

and the point of intersection of the ray is found by equating the derivatives of (5) to zero:

$$\eta = \partial V / \partial y, \quad \zeta = \partial V / \partial z, \quad \dots\dots(6)$$

which in our case reduces to $\eta = 3\beta(2r^2 + r^2 \cos 2\phi)$, $\zeta = 3\beta r^2 \sin 2\phi$. Therefore two rays from diametrically opposite points of a zone intersect in the same point η, ζ and must show interference. Substituting (6) into (5), the path difference with the principal ray becomes generally $V - y \partial V / \partial y - z \partial V / \partial z = V - 3V = -2V$, or twice this amount between the two interfering rays. Nienhuis finds that the observed appearance of the interference fringes is quantitatively explained in this way. In the lower part of the coma pattern, near the tip, the lower parts of large circles will overlap with the upper parts of much smaller ones. The two crossed systems of fringes give rise to the diamond pattern observed.

One further detail can better be illustrated on the astigmatism pattern. In this case the geometrical pattern is to be deduced in the same way from $V(y, z) = \beta r^2 \cos 2\phi = \beta(y^2 - z^2)$ giving $\eta = 2\beta y$, $\zeta = -2\beta z$, that is, all zones give concentric and proportional circles which are described in opposite directions. Therefore there is no overlapping here and we should expect an evenly illuminated circle instead of the observed four-pointed star. In the same way as above, we further obtain for the path difference in the receiving plane,

$$V - 2V = -V = -\beta(y^2 - z^2).$$

This means that our pattern, on interfering with an auxiliary spherical wave, should show the same hyperbolic fringes, with reversed sign, that would be found in the lens aperture in a Twyman interferometer. The method of the coherent background realizes this; figure 8 shows the result.

The four-pointed star may be explained by the diffraction at the diaphragm-edge. As is well known, the Fresnel diffraction at an edge of any form may be ascribed to rays emerging from the edge and spreading from the undiffracted ray only in directions perpendicular to the edge. In our case we must therefore

expect diffraction streaks issuing from each point on the circle of least confusion. As this circle is described in the opposite direction, the streaks will turn *against* the radius vector. The geometrical problem is easily solved, the streaks will envelop an asteroid, the equation to which will be $\eta^{\frac{1}{2}} + \zeta^{\frac{1}{2}} = (4\beta)^{\frac{1}{2}}$. This was tested experimentally by inserting a metal disc in the diaphragm opening, leaving free only a narrow annulus. The asteroid pattern then appeared, unobstructed by the intense geometrical pattern (figure 9(b)). The reverse is also possible: by throwing the shadow of a small circular opening on the lens aperture, the latter may be illuminated with an intensity decreasing towards the edge and vanishing at the edge itself. The edge effect was indeed absent in this case, only a simple circular pattern remaining. Such experimental tricks are not even necessary. It is found geometrically that the asteroid edge-pattern must remain unchanged, at least in form, when the focus is changed. It therefore comes out much clearer when the receiving plane is placed at one of the focal lines (figure 9(c)).

After these experiments the question arose: must we rest content with this solution of the problem? Evidently the problem is in essence a mathematical one; we have no reason to doubt the validity of the relatively simple diffraction integral, which was also the starting point for the development in case of small errors. The real difficulty was to find an *asymptotic expansion* for large wave numbers. Only a few months ago my assistant N. G. van Kampen attacked this problem again and found the solution. The "method of stationary phase" used goes back to Stokes and Kelvin. Its mathematical elaboration was well known to us, as it is due to my colleague J. G. van der Corput (1936). However, van Kampen had to extend it to two variables. Let me give a brief summary of the mathematical formulation. The problem is to develop an integral of the general form

$$\int_D kg(y, z)e^{ikf(y, z)} dy dz$$

into an asymptotic series for large values of the parameter k , the integral being extended over the domain D , the boundary of which consists of a finite number of analytical curves.

It is found that the ever increasing rapidity of fluctuation of the exponential causes the series to depend on the behaviour at a limited number of *decisive points*, which are of three kinds: (a) internal points at which the argument of the integrand is stationary, i.e. $\partial f/\partial y = \partial f/\partial z = 0$, (b) boundary points at which the argument is stationary along the boundary, i.e. $\partial f/\partial s = 0$, (c) corner points, or boundary points where two analytical curves join.

In the neighbourhood of any internal decisive point (y_0, z_0) , the exponential may be partially developed into a power series

$$\exp(ikf_0) \exp\{ik(\alpha_{11}y^2 + 2\alpha_{12}yz + \alpha_{22}z^2)\}\{1 + \dots ik\alpha_{31}y^3 + ik\alpha_{41}y^4\}$$

and this is integrated term by term between limits and $-\infty + \infty$. The resulting integrations are easily performed, they give a series with principal term

$$\pi g_0 e^{ikf_0} \left| \begin{array}{cc} \alpha_{11} & \alpha_{12} \\ \alpha_{12} & \alpha_{22} \end{array} \right|^{-\frac{1}{2}} \dots\dots (7)$$

with additional details about signs into which I shall not enter. The following terms are of order k^{-1} , k^{-2} etc. In a similar way, each decisive boundary point gives a series beginning with $k^{-\frac{1}{2}}$, $k^{-\frac{3}{2}}$ etc. and each corner point terms of order

k^{-1} , k^{-2} etc. The contributions of all decisive points must finally be added. Of course it is no easy matter to prove the mathematical validity of the whole procedure, but Professor van der Corput has just now succeeded in it.

Now, applying this to our diffraction integral, we may ordinarily put the amplitude g equal to one and $f = V(y, z) - y\eta - z\zeta$, and consider a circular boundary. The internal decisive points are then to be found from (6) and the principal term (7) is that of geometrical optics, with phases and interference taken into account. Even the intensity agrees exactly, the square of the last factor of (7) corresponding to the concentration of light rays by the curvature of the wave surface.

The second term comes from the boundary points, it corresponds to the beams diffracted by the edge. But here the theory gives more than the experimental treatment, which was not able to predict the intensity. Of course the new development also gives more than these two terms and these further terms could not be found in another way. We are now endeavouring to fill the gap between small and large errors by calculating some intermediate case, say astigmatism with $\beta = 10$, from both sides.

It is especially the general insight, however, which gains very much by the discovery of the asymptotic development. It shows that physical intuition combined with experimental ability may go far towards elucidating the main characteristics of phenomena, but that only an adequate mathematical treatment can give a satisfactory final solution.

REFERENCES

- BIJL, J., 1938, *Nieuw Arch. Wisk.* (2), **19**, 63.
 VAN CITTERT, P. H., 1934, *Physica*, **1**, 201.
 VAN DER CORPUT, J. G., 1936, *Compos. Math.*, **3**, 328.
 MICHELSON, A. A., 1890, *Phil. Mag.*, **30**, 1.
 NIJBOER, B. R. A., 1942, *Thesis, Groningen*.
 PEASE, F. G., 1931, *Ergebn. exakt. Naturw.*, **10**, 84.
 ZERNIKE, F., 1938, *Physica*, **5**, 785 ; 1946, *C.R. Réunions Opticiens*, in press.

The Structure and Growth of PbS Deposits on Rocksalt Substrates

BY A. J. ELLEMAN AND H. WILMAN

Applied Physical Chemistry Laboratory, Imperial College, London

MS. received 28 November 1947

ABSTRACT. The structure of PbS deposits condensed from the vapour *in vacuo* on to {001}, {110}, {111} and {443} rocksalt faces has been investigated by electron diffraction. The results suggest that the deposit atoms take up positions of least potential energy relative to the substrate, as far as is permitted by the disturbing effects of collisions of incident atoms with the initial deposit crystal nuclei, and by the limited surface mobility of the deposited atoms over the substrate. This view is also supported by the nature of the changes in crystal orientation which occur when initially random deposits are heated *in vacuo*.

§ 1. INTRODUCTION

LEAD sulphide is a semiconductor which has important applications owing to its photoconductive properties and especially its electron-emission sensitivity to infra-red radiation. Outlines of the properties of sensitive deposits, and of their production by sublimation *in vacuo* or low pressure oxygen have been published by Lee and Parker (1946), Starkiewicz *et al.* (1946), and Sosnowski *et al.* (1947 a, b).

A recent investigation (Wilman 1948) of such deposits by electron diffraction established that lanarkite, $\text{PbO} \cdot \text{PbSO}_4$, was formed as an oxidation product in the sensitization of the PbS by sublimation and baking in oxygen. With a view to studying further the form of the oxidation product and its connection with photoconductivity, single-crystal lead-sulphide deposits have now been prepared by condensation on rocksalt substrates. Rocksalt was chosen as likely to initiate a parallel orientation in the PbS because its axial length (5.639 Å.) is within 6% of that of PbS (5.929 Å.). The structure of these PbS deposits as shown by electron diffraction is described below. The results afford a picture of the mode of growth of a semiconductor condensed from its vapour on to an ionic substrate and provide a useful comparison with the growth of similarly prepared deposits of metals (Kirchner 1932, Lassen 1934, Lassen and Brück 1935, Rüdiger 1937, Wilman 1939).

§ 2. EXPERIMENTAL

Fresh rocksalt {001} faces about 8×5 mm. were prepared immediately before use by cleavage from an almost perfect natural crystal. The specimen was handled only with tweezers, and any stray fragments were lightly brushed off the face with a camel hair brush. Other faces used were ground on large cleavage blocks of rocksalt and smoothed on 0000 emery paper before lightly etching during a second or two in a fast stream of tap-water, followed by plunging immediately into absolute alcohol to stop further etching and to remove water and solution. Diffraction patterns of Kikuchi lines and spots elongated perpendicular to the shadow edge from such surfaces showed that the resulting smooth surfaces were free from random recrystallized NaCl.

The lead sulphide was prepared by precipitation from concentrated lead-nitrate solution by H_2S gas, followed by repeated washing with distilled water and drying in air at about 80°C . After driving off excess sulphur by heating *in vacuo*, a few milligrams were transferred to another pyrex tube and sublimed in a vacuum of the order of 10^{-3} mm. Hg on to rocksalt substrates.

The deposits were normally allowed to cool for at least five minutes before letting in air and transferring to a Finch-type electron-diffraction camera (Finch and Wilman 1937). A camera length of about 47 cm. and 50–65 kv. electrons were used. For transmission examination the rocksalt substrate was dissolved away in distilled water and the floating PbS film was picked up on a nickel gauze and dried *in vacuo* in the diffraction camera.

The rate of deposition and thickness of deposits were controlled roughly by eye to standard conditions, namely a rate of the order of 40 Å./sec. and thicknesses about 50, 250 or 1000 Å. (corresponding respectively to light yellow, translucent brown and opaque, or nearly so, by transmitted light). The thicknesses were estimated for a few cases by subliming known weights of PbS on to a measured area and visual comparisons with these were made for the remaining specimens.

§ 3. THE STRUCTURE OF THE DEPOSITS

The results are outlined in tables 1 and 2. The deposits prepared were mostly of medium thickness in view of the purpose for which they were later intended.

Deposits on apparently good ground and etched surfaces sometimes consisted of randomly disposed PbS crystals, even at the higher temperature range where strong orientation of the PbS was usually observed, all other conditions being kept constant. These results are not recorded in table 2 because they were attributed to the accidental presence of submicroscopic randomly disposed NaCl crystals on the main rocksalt substrate.

(i) *Orientations of the PbS*

In addition to the orientations listed in the tables, small proportions of randomly disposed PbS crystals were occasionally indicated by faint PbS rings in the diffraction pattern as in figure 2, but the orientation was usually very strong as shown by well-defined diffraction spot patterns, figures 1 and 6. In the case of deposits on {110} and {111} NaCl faces below about 150° c. the only orientation observed was with {111} planes parallel to the substrate with otherwise random azimuthal distribution.

(ii) *Twinning*

About one-third of the deposits on the {001} NaCl faces were found to possess strongly twinned structures on either {111} or {332} planes as shown by figure 2 and figures 3 and 4 respectively. The {332} twinning appears not to have been observed hitherto in the mineral galena. Deposits condensed from the vapour must of course grow from many nuclei formed initially on the substrate, not necessarily all in the same orientation; thus it is not clear whether these PbS crystals first grew in an orientation parallel to that of the NaCl and then twinned, or whether they were separate crystals in twin orientation. In the above cases, the crystals in the twin orientations to those which are parallel to the rocksalt are believed to arise actually by twinning from initially formed PbS crystals whose axes are parallel to those of the NaCl, because these latter crystals have in these cases developed faces (see (iii) below) which are parallel to the twinning plane. The planes in the {111} and {332} twin crystals, which would be junction planes with the NaCl {001} substrate, have lattice-point distributions which are only similar in spacing to the NaCl along one direction, i.e. the $\langle 110 \rangle$ rows of the NaCl, and growth in these orientations on the NaCl is, therefore, less likely than in the {001} orientation, though possible.

Diffraction spots due to {*hhl*} types of twinning are most prominently observed in the $\langle 110 \rangle$ azimuth, their positions being then obtained by the rotation of the main pattern (the stronger spots) about the central undeflected spot, through $2 \tan^{-1} l/h\sqrt{2}$, namely 70° 31' for {111} twinning (figure 2), 50° 28' for {332} twinning (figure 3). None of the seven deposits on {111} NaCl nor that on a {443} NaCl face showed any evidence of twinning, and only one on {110} NaCl showed a few extra spots not yet explained.

(iii) *Development of external crystal faces*

In most of the PbS deposits the crystals had tended to develop definite external faces to a greater or lesser extent as shown by elongation of the spots, especially those near to the shadow edge. In all cases the faces parallel to the substrate

Table 1. The structure of PbS deposits on {001} NaCl cleavage faces

Approx. temp. of substrate (° C.)	Approx. deposit thickness (A.)	No. of specimens with following characteristics									
		Total number	Orientations observed			Order of crystal diameter (A.)				Twinning observed	
			{001} †	{111} †	Random	L †		LR †			
						20-50	50-200	20-50	50-200		
20	50	6	5	—	1**	2-2**	2	2-2**	2	1	—
100-200	50	3	2 §	2 §	1	3	—	3	—	—	—
200-300	50	3	2+1 §	1 §	—	3	—	1	2	1	—
	250	11	10+1 §	1 §	—	4	7	2	9	1	4
	1000	3	3	—	—	1	2	—	3	3	—
	250*	3*	3	—	—	—	3	—	3	—	—
	1000*	3*	3	—	—	—	3	—	3	—	—
Approx. 400	250	2	2 §	2 §	—	—	2	—	2	—	—

* These deposits were kept at the deposition temperature for a further 30 sec. after conclusion of deposition before being allowed to cool in the normal manner.

† Orientation with (001) planes parallel to {001} cleavage surface of the NaCl with the cubic axes parallel to those of NaCl.

‡ This denotes here {111} PbS parallel to {001} NaCl cleavage face and <110> lattice row parallel to <110> of NaCl.

§ These specimens showed both orientations together, figure 5.

¶ L and LR are reckoned with respect to the substrate surface.

** These specimens had a mean crystal diameter < 20 A.

Table 2. The structure of PbS deposits on {110}, {111} and {443} NaCl faces

Approx. temp. of substrate (° C.)	Approx. deposit thickness (Å.)	No. of specimens with following characteristics					
		Total number	Orientations observed	Order of crystal diameters (Å.)			
				L†		⊥R†	
				20-50	50-200	20-50	50-200
<i>Deposits on {110} NaCl face</i>							
20	50	1	Slight {111} L to substrate	1	—	1	—
100-200	50	1	Random	—	1	—	1
200-300	50	1	L* 1	1	—	1	—
	250	2	L* 2	1	1	1	1
<i>Deposits on {111} NaCl face</i>							
100-200	50	1	Slight {111} L substrate	1	—	1	—
	1000	1	Random	—	1	—	1
200-300	50	4	L* 4	4	—	4	—
	250	2	L* 2	2	—	2	—
<i>Deposit on {443} NaCl face</i>							
200-300	250	1	L* 1	1	—	1	—

* ||L denotes PbS cubic axes parallel to those of the NaCl.

† ||L and ⊥R are reckoned with respect to the substrate surface.

were the most prominent and were often very extensive and atomically smooth as shown by the degeneration of the spot pattern into almost continuous parallel lines perpendicular to the shadow edge. This is shown in figure 1, characteristic of those deposits which were kept at the deposition temperature (200-300° C.) for a further 30 sec. after deposition. The larger size of the crystals in these deposits was shown by the appearance of diffuse Kikuchi bands, as well as by the sharpness of the vertical streaks.

The patterns from one of the {332}-twinned deposits on {001} NaCl in the $\langle 100 \rangle$ and the $\langle 110 \rangle$ azimuths showed that the crystals in the main orientation parallel to the NaCl had developed {100}, {110}, {120}, and {335} faces, the streaks in the pattern corresponding to {335} faces making an angle of 40° with the plane of incidence in figure 3. The elongation of the twin spots perpendicular to the shadow edge showed that the twin lattices also had developed boundary faces parallel to the substrate and therefore of {335} type with respect to the cubic axes of the twin lattices. In another twinned PbS deposit having a smaller proportion of {332} twinning there were strongly developed {001} faces, and {120} faces were prominent though {110} and {335} were not observed.

In PbS deposits on {110} and {111} NaCl faces there was slight elongation of spots near and normal to the shadow edge, showing a tendency of the crystals

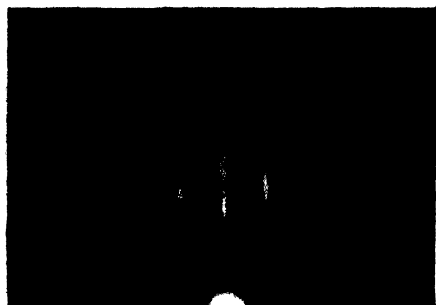


Figure 1. PbS on {001} NaCl; {001} orientation, $\langle 110 \rangle$ Az.



Figure 2. PbS on {001} NaCl; {001}, $\langle 110 \rangle$ Az., {111} twinning.

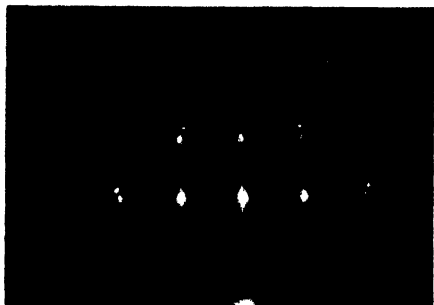


Figure 3. PbS on {001} NaCl; {001}, $\langle 100 \rangle$ Az., {332} twinning.

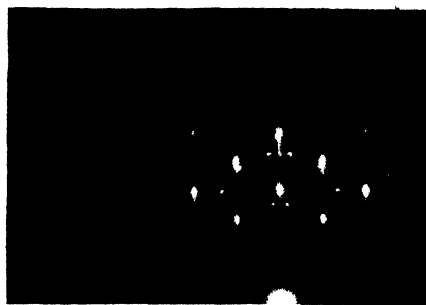


Figure 4. PbS on {001} NaCl; {001}, $\langle 110 \rangle$ Az., {332} twinning.

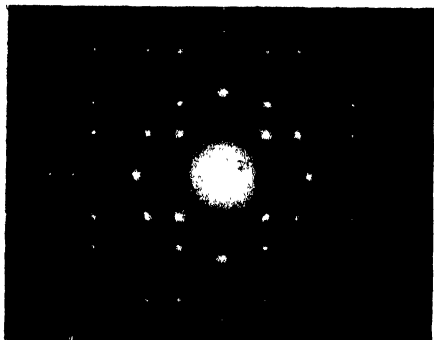


Figure 5. PbS grown on {001} NaCl. {001} and {111} orientations.

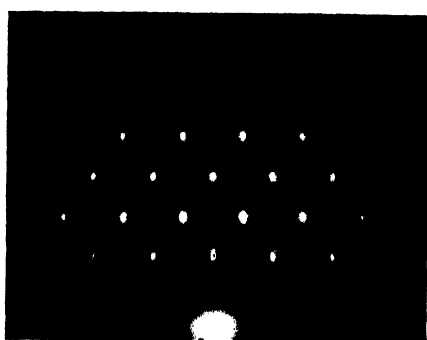


Figure 6. PbS on {110} NaCl; parallel orientation.

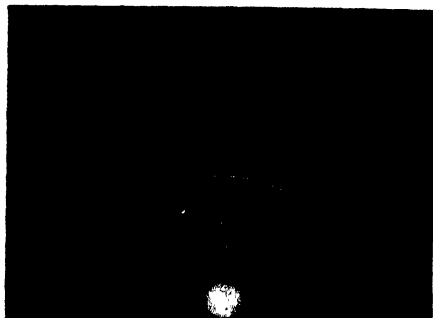


Figure 7. Random PbS on {111} NaCl.

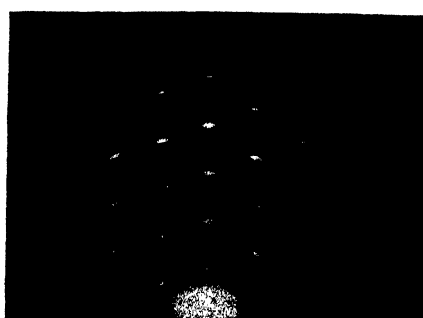


Figure 8. Deposit of Figure 7 heated in vacuum to $\sim 250^\circ\text{C}$.

in the deposit to form faces parallel to the substrate. In two specimens on {110} NaCl faces there was pronounced development of {001} boundary faces, inclined at 45° to the substrate, and the large extent of the spot pattern is also consistent with the development of such small projections on the deposit surface.

The thin PbS deposit prepared on a {443} NaCl face had all its crystals very strongly oriented with their cubic axes parallel to those of the NaCl. The spot patterns were as free from arcing as those of figures 3 and 4, though without any elongations representing definite boundary faces. The pattern previously obtained from the {443} NaCl face showed Kikuchi bands and relatively few but strong spots elongated normal to the shadow edge, showing almost perfect {443} boundary faces.

(iv) Lattice constants

In two cases of PbS deposits at $200\text{--}300^\circ\text{C.}$ on {001} NaCl faces, the lattice constants of the PbS were measured from the patterns obtained by transmission after dissolving away the NaCl. In one case the reference material was graphite ($d_{110}=1.230\text{ \AA.}$) giving $a_{\text{PbS}}=5.931\text{ \AA.}$, and in the other case NaCl crystallized from solution ($d_{200}=2.820\text{ \AA.}$) giving $a_{\text{PbS}}=5.927\text{ \AA.}$ These values are close to the average of $a=5.929\text{ \AA.}$ previously found for PbS layers formed by passing H_2S gas over lead-nitrate or lead-acetate solution or sublimation of such material *in vacuo* in pyrex cells (Wilman 1948).

§ 4. THE EFFECT OF HEATING PbS DEPOSITS ON NaCl SUBSTRATES *IN VACUO*

It has been mentioned above that the continued heating of PbS deposits *in vacuo* at $200\text{--}300^\circ\text{C.}$ after deposition on the {001} NaCl faces, only resulted in elimination of twinned structure, increase of crystal diameter and strong development of faces parallel to the substrate. No change of the main orientation was observed. It seemed possible, however, that (i) deposits consisting of random crystals as a result of low mobility of the deposit atoms on the substrate at the temperature of deposition might develop a two-degree orientation relative to the substrate lattice when heated *in vacuo*; and (ii) that the rate and kind of orientation developed should depend on the temperature at which the heat treatment is carried out and on the distribution of atoms or ions in the substrate surface.

To test this a 1000 $\text{\AA.}$ thick PbS layer was deposited at $100\text{--}150^\circ\text{C.}$ on a {111} NaCl face and was found to consist of random crystals about 120 $\text{\AA.}$ in diameter. When this was heated for six minutes at $200\text{--}300^\circ\text{C.}$ *in vacuo* it developed strong single-crystal structure (figure 8) with a {111} plane parallel to the {111} substrate surface and the cubic axes parallel to those of the NaCl. A further confirmation was obtained in the case of a 50 $\text{\AA.}$ thick random layer deposited at room temperature on a {001} NaCl cleavage face. Though initially composed of random crystals of only 10–20 $\text{\AA.}$ mean diameter, on heating to $200\text{--}300^\circ\text{C.}$ *in vacuo* for four minutes it developed strong two-degree {001} and {111} orientation similar to that in layers deposited at this temperature on {001} NaCl. In both cases there was a large and rapid increase in crystal size at this temperature of heating, to a mean diameter of at least 300 $\text{\AA.}$ though less than 500 $\text{\AA.}$

§ 5. ORIENTATION OF NaCl ON PbS

Several cases were observed, e.g. figure 6, where traces of NaCl had evidently been present on the walls of the pyrex tube and has sublimed on to the already

deposited and oriented PbS. These NaCl deposits were found in all cases to have the same orientation as the underlying PbS deposit, whether this was oriented or random.

§ 6. DISCUSSION

(i) *The orientations observed*

The orientations of the PbS crystals relative to the {001}, {110}, {111} and {443} NaCl substrates agree with the conditions generally observed, that in such cases of strongly oriented overgrowths there is a similarity of periodicity in at least one set of parallel lattice rows in the deposit and substrate crystal. The development of {111} orientations on {001} NaCl at temperatures above about 250° C. as well as the first-developed {001} orientations is analogous to the case of silver deposits formed on {001} NaCl faces over a similar range of temperatures (Brück 1936). On {111} and {443} NaCl faces, the consistent growth of PbS with cubic axes parallel to those of the NaCl, and not in more than one orientation geometrically equivalent relative to the surface plane, is analogous to the continued growth of NaCl on NaCl, building up a continuous pattern of positive and negative ions. Thus it would seem that not only the surface atoms govern the orientation in the growth of such ionic crystals but also those in the underlying net-plane.

(ii) *Twinning*

The proportion of specimens having a {111} or {332} twinned structure is only about 30% compared with almost 100% for those of silver. This may perhaps be associated with the smaller difference of lattice constants of the substrate and PbS deposit (cf. Wilman 1939). On the other hand, it seems likely that here the PbS crystals grow initially in an orientation parallel to the NaCl and twin during their further growth, whereas silver is strongly {111} twinned in very thin layers; but the crystals in the four twin orientations grow together (Menzer 1938), so that surface regions of thick silver layers are formed of crystals whose axes are parallel to those of the NaCl.

(iii) *Lattice constants*

The present estimations from transmission patterns show more accurately than was previously possible from reflection patterns (Wilman 1948) the striking constancy of the lattice axial length, 5.929 Å., of PbS prepared both by chemical deposition and by sublimation *in vacuo*.

(iv) *Interpretation of deposit structure in terms of a hypothesis of lowest potential energy state modified by surface mobility and rate of deposition of atoms from the vapour stream*

It is now clear that oriented deposits must arise through the initially deposited atoms taking up stable positions where they have lowest potential energy on substrate crystal surfaces which consist of a periodic distribution of atoms or ions. Similar views have been expressed for example by Stranski (1928). In like manner further atoms take up packing positions having least potential energy, and thus consistent with the normal deposit crystal structure and atomic distribution, i.e. orientation, in the first layer. The attainment of such a state is, however, dependent especially upon the rate of deposition and on the thermal movements and the strength of forces between the deposit and substrate ions compared with those of the deposit and substrate alone. Increase in temperature

of the substrate, by increasing the amplitude of thermal vibrations of the substrate surface atoms, should assist the growth of strongly oriented deposits, and this has indeed been confirmed by many experiments including the present results on PbS.

The effect of rate of deposition on the number and size of the deposit crystals at any stage of the deposition on single crystal or amorphous substrates has not been quantitatively explored in the present experiments on PbS, but the general observations so far made may be explained as follows: The atoms first falling on the substrate have a certain degree of mobility over the surface before losing most of the kinetic energy which they possessed on arrival; thus, sooner or later, they meet other atoms and aggregate to form small nuclei. The greater the number of atoms incident on unit surface area per unit time, the larger will be the number of these nuclei formed per $\text{cm}^2/\text{sec.}$ and the smaller their lateral extent, and also the smaller the distances between these nuclei on the still uncovered substrate. The higher the rate of arrival of the atoms per $\text{cm}^2/\text{sec.}$, the greater is the number of atoms bombarding each nucleus per sec., and these will therefore hinder the development of the orientation which would be imposed by the substrate and result in a correspondingly weaker orientation or even a completely random crystal deposition. This picture is analogous to the Brownian movement of small particles in a liquid due to bombardment by the atoms of the liquid surrounding them.

The above experiments on PbS show that, even with a relatively high rate of deposition of about 40 Å. thickness per second, strongly oriented PbS deposits on NaCl were obtained at substrate temperatures of about 150°C. As in the case of silver and other metals under similar conditions, random or weakly oriented PbS deposits tend to be formed at or near room temperature; and no strong development of orientation with increase of thickness was observed, such as occurs when the crystals develop large plane faces perpendicular to the incident beam direction. (Kirchner 1932, Burgers and Dippel 1934, Beeching 1936, Burgers and van Amstel 1936). These results, together with the observation that random-crystalline PbS deposits do not develop orientation unless heated well above room temperature, suggest that the PbS molecules have low mobility over NaCl surfaces at room temperature, but that the mobility rapidly increases with temperature.

The PbS deposits prepared at room temperature tended to be orientated apparently because the substrate surface temperature became more or less raised by absorption of radiant energy from the hot test-tube wall from which the PbS was being sublimed. The temperature of the substrate at its surface is difficult either to determine or control. It is a result of equilibrium between energy falling on to the surface by radiation from the oven—or filament—source of atoms, and energy leaving the surface by re-radiation and by conduction to the lower layers of the substrate. It was found that PbS deposits about 50 Å. thick on an NaCl-crystal substrate 2 to 3 mm. thick were often weakly oriented, but when 10 mm. thick NaCl crystals of about the same area were used, the 50 Å. thick PbS deposits obtained were of quite randomly disposed crystals only 10–20 Å. in diameter. This demonstrates the need for using substrates of high thermal capacity, if necessary backed by a metal mass of high conductivity, to ensure that the substrate surface temperature is not much higher than that of the rest of the substrate.

(v) *Recrystallization of initially random PbS deposits*

According to the above view of the dependence of crystal growth on the deposit conditions, we would expect that even if the crystals of the deposit are initially randomly disposed, subsequent heat treatment should enable rearrangement of the atoms to take place in such a way that the state of least potential energy with the substrate, consistent with the surface mobility, is attained. This state should be the same as that which would have been obtained if the layer had been deposited at this temperature. The results of the experiments in §4 do, in fact, lend further support to the hypothesis that the lowest potential energy state tends to be attained as far as the limited surface mobility of the deposit atoms permits. In this connection it is noteworthy that both {001} and {111} orientations of the PbS were developed together on the {001} NaCl face when heated *in vacuo* to 200–300° C.

The fact that random crystalline (or amorphous) deposits on heating take up an orientation characteristic of the substrate and its surface temperature, which under these conditions becomes that of the heating enclosure, enables the relation to be studied between the orientation of thin deposits and the nature and surface temperature of the substrate, without the uncertainty of measurement of the actual surface temperature during the deposition process. This should be especially helpful in the case of deposit materials which show a variety of orientations within a relatively narrow temperature range.

§7. SUMMARY

The structure of PbS deposits condensed from the vapour *in vacuo* on to {001}, {110}, {111} and {443} rocksalt faces has been investigated by electron diffraction. The results obtained agree with the theory that the deposit atoms take up positions of least potential energy relative to the substrate, as far as is permitted by the disturbing effects of collisions of incident atoms with the initial deposit crystal nuclei and by the limited surface mobility of the deposited atoms over the substrate. Further confirmation of our theory is afforded by changes of orientation of initially random deposits when heated *in vacuo*, and lead to the conclusion that such experiments should enable a more precise investigation to be made of the relation between the deposit orientation and the temperature of the substrate surface.

ACKNOWLEDGMENTS

This work forms part of a study of the fundamental mechanism of crystal growth being carried out in this Laboratory and it is a pleasure to thank Professor G. I. Finch for his interest and encouragement.

A. J. Elleman also thanks the Department of Scientific and Industrial Research for a grant which enabled him to devote his full time to this research.

REFERENCES

- BEECHING, R., 1936, *Phil. Mag.*, **22**, 938.
BRÜCK, L., 1936, *Ann. Phys. Lpz.*, **26**, 233.
BURGERS, W. G., and VAN AMSTEL, J. J. A., 1936, *Physica*, **3**, 1057.
BURGERS, W. G., and DIPPEL, D. J., 1934, *Physica*, **1**, 549.
FINCH, G. I., and WILMAN, H., 1937, *Ergebn. exakt. Naturw.*, **16**, 353.
KIRCHNER, F., 1932, *Z. Phys.*, **76**, 9, 576.
LASSEN, H., 1934, *Phys. Z.*, **35**, 172.
LASSEN, H., and BRÜCK, L., 1935, *Ann. Phys., Lpz.*, **22**, 65.
LEE, E., and PARKER, R. C., 1946, *Nature, Lond.*, **158**, 518.

- MENZER, G., 1938, *Naturwissenschaften*, **26**, 385 ; *Z. Kristallogr.*, **99**, 378 and 410.
 RÜDIGER, O., 1937, *Ann. Phys., Lpz.*, **30**, 505.
 SOSNOWSKI, L., SOOLE, B. W., and STARKIEWICZ, J., 1947, *Nature, Lond.*, **160**, 471.
 SOSNOWSKI, L., STARKIEWICZ, J., and SIMPSON, O., 1947, *Nature, Lond.*, **159**, 818.
 STARKIEWICZ, J., SOSNOWSKI, L., and SIMPSON, O., 1946, *Nature, Lond.*, **158**, 28.
 STRANSKI, I., 1928, *Z. phys. Chem.*, **136**, 259.
 WILMAN, H., 1939, *Proc. Phys. Soc.*, **51**, 625 ; 1948, *Ibid.*, **60**, 341.

A Determination of the Ratio of the Masses of π - and μ -Mesons by the Method of Grain-counting *

By C. M. G. LATTES, G. P. S. OCCHIALINI AND C. F. POWELL
 H. H. Wills Physical Laboratory, University of Bristol

MS. received 10 April 1948

§ 1. INTRODUCTION

IN a recent article (Lattes 1947), we gave an account of a determination, by the method of grain-counting, of the ratio of the masses of the π - and μ -mesons observed in photographic emulsions. We have found more than thirty events in which the secondary μ -meson produced by a π -meson at rest comes to the end of its range in our emulsions. This new material, of which the main features are summarized in table 1, allows us to increase the accuracy of the determination,

Table 1. Ranges of π - and μ -mesons in individual examples of the μ -decay

Event No.	Ranges of		E_μ (mev.)	Event No.	Ranges of		E_μ (mev.)
	π -meson (microns)	μ -meson (microns)			π -meson (microns)	μ -meson (microns)	
1	134	613	4.08	17	477	585	3.97
2	1287	615	4.09	18	12	685	4.36
3	124	591	4.00	19	52	644	4.21
4	106	633	4.16	20	62	594	4.00
5	453	615	4.09	21	230	570	3.91
6	105	567	3.91	22	115	570	3.91
7	906	606	4.05	23	488	680	4.34
8	248	627	4.13	24	96	667	4.29
9	231	648	4.22	25	409	626	4.13
10	81	590	3.99	26	91	623	4.12
11	57	609	4.06	27	990	602	4.04
12	634	567	3.90	28	92	619	4.10
13	48	611	4.06	29	15	598	4.01
14	1297	603	4.04	30	91	639	4.18
15	72	591	4.00	31	2290	660	4.27
16	57	590	3.99				

Mean values : $R_\mu=614 \mu$; $E_\mu=4.08$ mev.

and in this article we describe in more detail the method of making the measurements and the results which follow from them. We find that the mass ratio $m_\pi/m_\mu=1.65 \pm 0.11$, the limits corresponding to the "probable error" arising

* An account of the main features of a number of lectures given in Copenhagen in December 1947, in Zurich in January 1948, and in a number of colloquia in English Universities.

from statistical fluctuations alone. It is reasonable to assume that the μ -mesons appear as a result of a spontaneous decay of the heavier π -mesons, and that the former are identical in type with the ordinary cosmic-ray mesons of mass $200 m_e$. It then follows from the present results that the momentum balance in the decay is provided by a neutral particle, a neutral meson with a rest-mass equal to $115 \pm 30 m_e$, where m_e is the mass of the electron. The observations also allow us to draw conclusions about the decay constants of the different particles, and we have thus determined the following limits for the half-value periods: $\tau_\pi > 0.4 \times 10^{-10}$ sec., $\tau_\mu > 2 \times 10^{-10}$ sec., and $\tau_\nu > 2.6 \times 10^{-21}$ sec., where τ_ν is the half-value period for the postulated neutral meson.

§ 2. PRINCIPLE OF THE METHOD

In determining the quantity m_π/m_μ by grain-counts, we make the usual assumption that a particle with a charge $|e|$, energy E Mev., and range in the emulsion R microns, loses energy at a rate which depends only on its velocity, v , and is independent of its mass, m : $\partial E/\partial R = f_1(v)$. It is then easily shown that $\partial E/\partial R = f_2(R/m)$ and $E = m\phi(R/m)$.

We assume, secondly, that if $N(R)$ is the total number of grains in the track of a particle of range R microns, the number of grains per micron, $\partial N/\partial R$, in an element of length of the trajectory, is a function only of the rate of loss of energy, $\partial E/\partial R$, of the particle. It follows that

$$N = mF(R/m). \quad \dots\dots(1)$$

Equation (1) allows us, in principle, to determine the ratio m_π/m_μ from the observed numbers of grains in the track of π - and μ -mesons.

In making grain-counts, we employ a $\times 95$ "Achromatic" objective with $\times 10$ eyepieces, and count the numbers of grains in successive intervals along the trajectory of length 50μ . These intervals are measured along the projection of the track on the plane of the emulsion. With this procedure we shall tend to over-estimate the grain-density in a track which dips relative to the plane of the emulsion, through "fore-shortening". We have therefore confined our measurements to those events in which the angle of dip, β , at every point in the trajectories of the primary and secondary particles is never greater than 15° . The errors due to "fore-shortening" are then small.

A second source of error exists owing to the variation of the degree of development of an emulsion with depth which occurs when a plate is processed by conventional methods. This has the consequence that the grain density may be different in the tracks of two particles produced contemporaneously, in regions of equal ionization, if they are situated at different depths in the emulsion.

Thirdly, as we pointed out in a previous communication, there is not a completely random distribution of grains in the gelatine in the emulsions at present available. "Islands" appear to exist in which the concentration of silver-halide grains is significantly higher, or significantly lower, than we should expect to find as a result of random fluctuations alone. This feature seems to be particularly noticeable in boron-loaded plates. The measurements show, however, that the influence of this factor on the total numbers of grains in the relatively long tracks on which our observations are made is not more serious than the purely statistical fluctuations associated with the finite number of grains in a track.

All the above sources of error will effect the measurements on π - and μ -mesons indiscriminately; and the errors they produce in the determination of the mean

value of the mass-ratio will be small if the measurements are made on a sufficiently large number of events. They contribute, however, to the variation among the individual values of the mass-ratio.

The following difficulties cannot at present be avoided. First, at the end of the range of a particle the ionization is sufficiently intense to produce an almost continuous succession of grains which cannot be individually resolved. We refer to this effect as "clogging". No accurate counts can be made in this region of a track, and we therefore adopt the following convention. We define the number of grains in a segment of the trajectory of length r , in which there is a contiguous succession of grains, as equal to r/a , where a is the mean "diameter" of the grains. In meson tracks "clogging" is serious only in the last 50μ from the end of the range, and for such short tracks no reliable mass determinations are possible.

Secondly, when the rate of loss of energy of the particle is very small, the grain density becomes so low that the counts become unreliable because of the general "background" of grains. This sometimes makes it difficult to decide whether a particular grain was produced by the passage of the particle or is due to "fog". In our emulsion this difficulty begins to become appreciable at 400μ and 650μ , respectively, from the ends of the ranges of μ - and π -mesons.

§ 3. CONSISTENCY OF GRAIN-COUNTS

Even in those parts of the trajectory of a particle in which "clogging" is not serious, the individual grains cannot always be completely resolved. It is therefore necessary to adopt a convention that a particular size of grain, for example, shall be regarded as two, that very small grains shall be neglected, etc. We have found in practice that if different observers establish a common convention, by reference to suitable photo-micrographs of characteristic tracks, it is possible to obtain the following degree of consistency between different measurements. An experienced observer can determine the total number of grains in the last 400μ of the track of a particular meson with a consistency of 3%, and there is a similar degree of consistency between the observations of two different observers. In practice we have made a succession of five determinations of the number of grains, N_π and N_μ , in each of the tracks of the π - and μ -mesons of a particular event, making counts on the two tracks alternately. The ratio of the numbers, N_π/N_μ , on which the mass ratio depends, is then found to be consistent to about 3%. Because of the finite number of grains in a track, the statistical fluctuation in the numbers N_π or N_μ to be expected in going from one event to another is of the order of 6%, and the difference in the average number of grains in equal lengths of the tracks of π - and μ -mesons about 13%.

§ 4. RESULTS

In table 2 we show the results of grain counts on those events which satisfy the conditions that at no point in either of the trajectories is the angle of dip greater than 15° . The table shows the number of grains, $N(R)$, in a residual range R microns—for intervals 50μ long—up to a maximum of 400μ for μ -mesons and of 650μ for π -mesons. The numbers in each vertical column correspond to the measurements on a particular event, so that the results for the two types of mesons can be compared.

It will be seen that the corresponding results for different events are subject to considerably greater fluctuations than those corresponding to statistical effects

Table 2. Grain counts on the track of π^- and μ^- -mesons

Event No.	2	14	7	12	5	17	8	9	1	4	3
Range (microns)	No. of grains in tracks of										
	μ^- -mesons										
50	66	55	54	72	61	60	60	66	45	53	69
100	109	97	90	127	105	104	105	116	80	91	118
150	144	125	111	166	147	133	146	164	97	125	162
200	179	157	132	204	180	162	183	205	119	154	205
250	211	184	157	235	216	198	216	242	142	179	249
300	236	201	184	266	239	233	243	276	167	196	287
350	265	216	201	287	266	261	274	319	188	224	319
400	289	233	220	312	293	285	301	341	204	251	339
	π^- -mesons										
50	75	57	66	72	68	65	71	80	41	58	75
100	129	96	114	120	121	108	126	142	81	101	129
150	171	136	151	159	179	154	175	194	111	140	
200	218	172	180	198	222	196	214	235	135		
250	260	200	208	234	250	222	258				
300	298	231	233	266	278	253					
350	330	257	243	300	307	286					
400	363	282	268	332	340	313					
450	391	306	292	367	367						
500	416	329	313	400							
550	451	355	332	432							
600	478	376	343	466							
650	508	392	359								

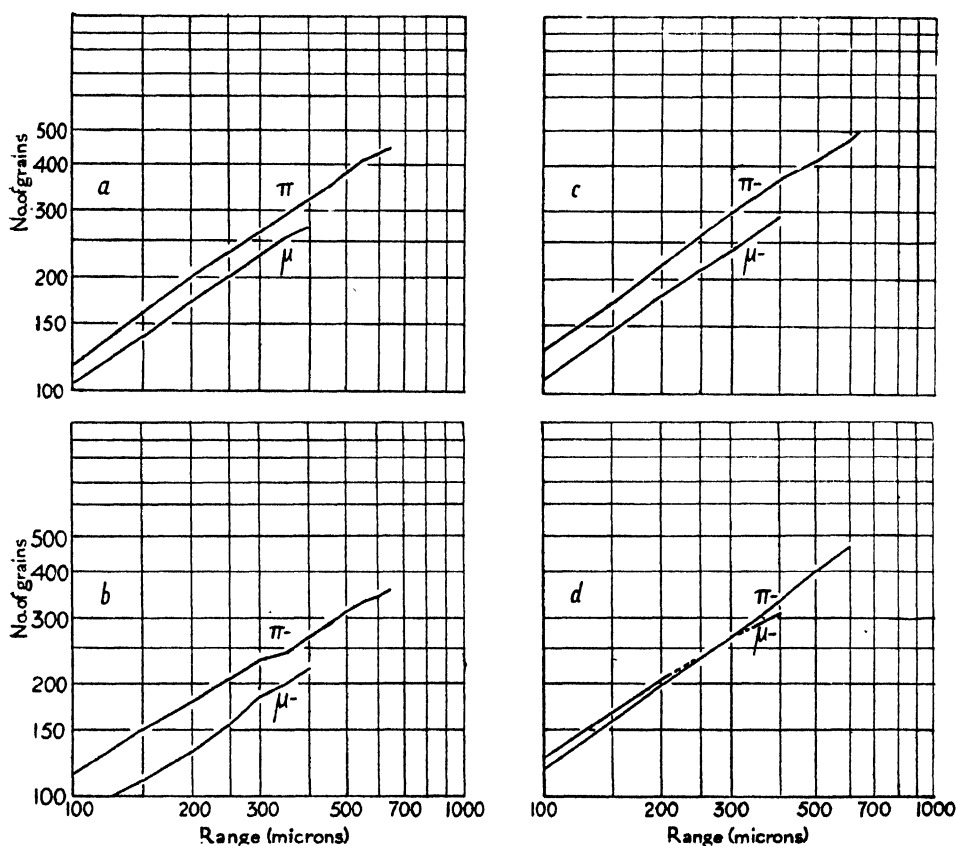


Figure 1. Logarithmic plots of the number of grains in the tracks of π^- and μ^- -mesons for four examples of the μ^- -decay. (d) shows the results for the event which displays the least difference in the number of grains in the tracks of the associated π^- and μ^- -mesons.

alone. We attribute these variations to "fading" and to the other factors discussed above. The results for four typical events are represented in figure 1, in which $\log N$ is plotted against $\log R$. Figure 1(d) corresponds to the event which displays the least difference between the results for the π - and μ -mesons.

In order to compare the results for the different events, we normalize the values for the different observations. We first determine a normalizing factor for each event, k_i , defined as the ratio of the average number of grains in the last 400 μ of the different μ -mesons, $N_\mu^m(400)$, to the number of grains, $N_\mu^i(400)$, in the same length of the particular μ -meson under consideration: $k_i = N_\mu^m(400)/N_\mu^i(400)$. Table 3 shows the values $\mathcal{N}(R)$ obtained by multiplying the numbers $N(R)$ for

Table 3. Values of $\mathcal{N}(R)$, for π - and μ -mesons, after normalizing to correspond to a constant value of $\mathcal{N}_\mu(400)$

Event No.	2	14	7	12	5	17	8	9	1	4	3	$\mathcal{N}^m(R)$
k_i	0.956	1.197	1.268	0.894	0.952	0.979	0.927	0.818	1.368	1.112	0.823	
Range (microns)												
	μ -mesons											
50	64	66	69	64	58	59	56	54	62	59	57	60 ± 2
100	105	116	114	114	100	102	97	95	110	101	97	104 ± 2
150	139	150	141	148	140	130	135	134	133	139	133	138 ± 2
200	173	188	167	182	171	159	170	168	163	171	169	171 ± 3
250	204	219	199	210	206	194	200	198	194	199	205	202 ± 3
300	228	241	233	238	228	228	225	226	228	218	236	230 ± 3
350	256	259	255	257	253	256	254	261	257	249	263	256 ± 3
400	279	279	279	279	279	279	279	279	279	279	279	279 ± 3
	π -mesons											
50	72	68	84	64	65	64	66	65	56	64	62	66 ± 2
100	124	115	145	107	115	106	117	116	111	113	106	116 ± 2
150	165	163	191	148	170	151	162	159	152	156		160 ± 3
200	210	206	228	177	211	192	198	192	185			198 ± 3
250	251	239	264	209	238	217	239					232 ± 4
300	288	277	295	238	265	248						263 ± 4
350	318	308	308	268	292	280						290 ± 5
400	350	338	340	297	324	306						320 ± 5
450	377	366	370	328	349							348 ± 6
500	401	394	397	358								375 ± 6
550	435	425	421	386								404 ± 7
600	461	450	435	417								426 ± 7
650	490	469	455									449 ± 8

The values of $\mathcal{N}_\pi^m(R)$ for π -mesons have been obtained by determining the average number of grains, $\Delta \mathcal{N}_\pi^m(R)$, in successive intervals of length 50 μ , and summing over the total range R .

the different events given in table 2, by the appropriate normalizing factor k_i , i.e. $\mathcal{N}(R) = k_i N(R)$. It may be emphasized that the mass ratio m_π/m_μ depends only on N_π/N_μ , and the values deduced from the observations on the individual events will not be changed by the use of the normalizing factor. Table 3 also contains the mean values for the number of grains, $\mathcal{N}_\pi^m(R)$, in a range R microns. The results are represented diagrammatically in figure 2, in which values of $\log \mathcal{N}^m(R)$ are plotted against $\log R$, for both π -mesons (upper line) and μ -mesons (lower line).

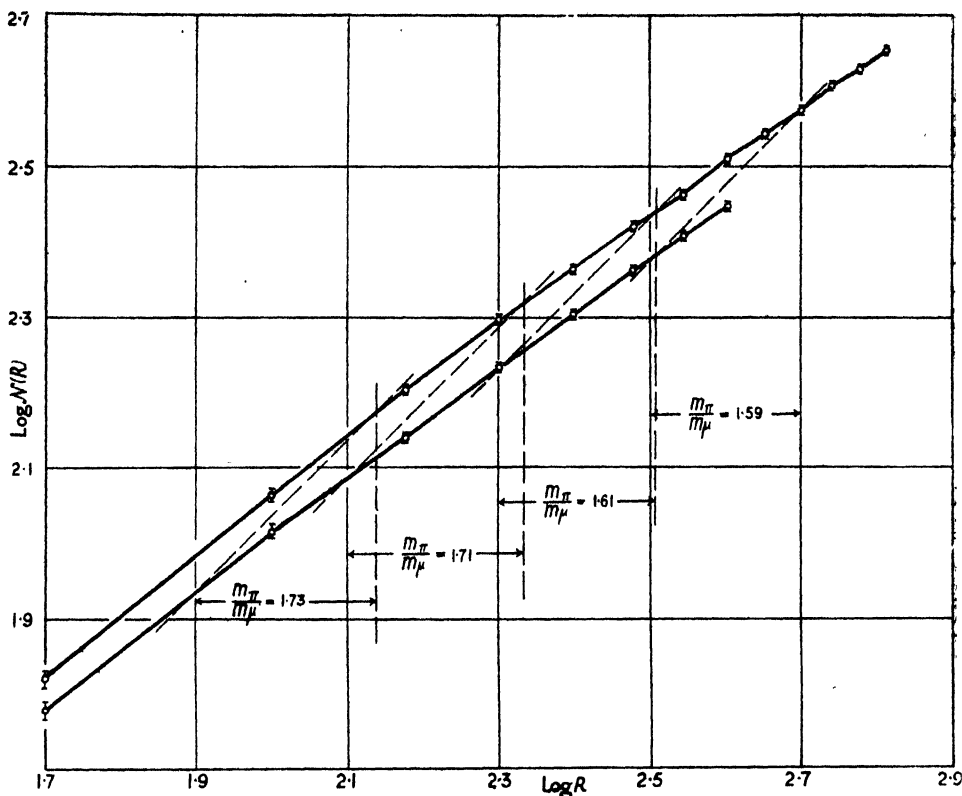


Figure 2. Plots of $\log \mathcal{N}_\pi(R)$ and $\log \mathcal{N}_\mu(R)$ against $\log R$. The upper and lower curves are drawn through the results for π - and μ -mesons respectively.

§ 5. DETERMINATION OF m_π/m_μ

The determination of the quantity m_π/m_μ can be carried out by any one of the following methods, the first of which was suggested to us by D. H. Perkins. It may be emphasized that they represent merely alternative modes of analysis, and not independent determinations.

(a) Perkins' method

It follows from the equation $N = mF(R/m)$ that if we establish a graph showing the relationship between $\mathcal{N}_\mu^m(R)$ and R for μ -mesons, we can deduce the corresponding curve for π -mesons for any assumed value of the mass-ratio m_π/m_μ . For each point $(\mathcal{N}_\mu^m(R), R)$ in the graph for μ -mesons we construct the corresponding point $(\mathcal{N}_\mu^m(R)m_\pi/m_\mu, Rm_\pi/m_\mu)$. The locus of these points then corresponds to the relation between $\mathcal{N}_\pi^m(R)$ and R for π -mesons. Alternatively, if we plot the relation between $\log \mathcal{N}_\mu^m(R)$ and $\log R$, the corresponding values of $\log \mathcal{N}_\pi^m(R)$ and $\log R$ are given by a simple translation of magnitude $\log(m_\pi/m_\mu)$ in both coordinates.

Conversely, if the relations between $(\log \mathcal{N}_\pi^m(R), \log R)$ and $(\log \mathcal{N}_\mu^m(R), \log R)$ have been established experimentally, values of m_π/m_μ can be deduced by drawing a succession of lines at 45° to the axes and measuring the translations, AB in figure 2, to which they correspond. The values so obtained are shown in table 4.

Table 4. Values of m_π/m_μ deduced from normalized grain-counts

Range (microns)	50	100	150	200	250	300	350	400
m_π/m_μ	1.70	1.67	1.73	1.63	1.58	1.55	1.60	1.64
"Probable error" ($\pm$)	0.18	0.14	0.13	0.13	0.13	0.12	0.12	0.11

The various values are not independent, and that with the greatest statistical weight is 1.64 ± 0.11 .

The values of m_π/m_μ for the different values of R_μ are not independent, and that with greatest statistical weight is $m_\pi/m_\mu = 1.64 \pm 0.12$, the limits corresponding to the magnitude of the "probable error".

(b) *Determination of the form of the function $N = mF(R/m)$*

It will be seen from figure 2 that the relationships between $\log \mathcal{N}_\pi^m(R)$ and $\log R$ and between $\log \mathcal{N}_\mu^m(R)$ and $\log R$ are linear for values of R greater than 100μ to within the standard deviations of the measurements. We can therefore write

$$\mathcal{N} = km(R/m)^n \quad \text{or} \quad \mathcal{N} = km^{-n}R^n. \quad \dots\dots(2)$$

We have determined the value of n from the best straight line drawn through the points representing the measurements on π - and μ -mesons in figure 3, and have obtained the values $n_\pi = 0.701 \pm 0.014$, $n_\mu = 0.721 \pm 0.014$, and the mean value $n = 0.711 \pm 0.011$. Equation (2) can therefore be written $\mathcal{N}^m = km^{0.289}R^{0.711}$. From the position of the straight lines in figure 2 we find that $km_\mu^{0.289} = 3.95 \pm 0.05$ and $km_\pi^{0.289} = 4.53 \pm 0.07$, so that $\mathcal{N}_\pi^m = 4.53R^{0.711}$ and $\mathcal{N}_\mu^m = 3.95R^{0.711}$. The values of $\mathcal{N}_\pi^m$ and $\mathcal{N}_\mu^m$ calculated from this formula for different values of R show no significant difference from those based on the grain-counts given in table 3, except for values of R less than 100μ .

Alternatively we can plot the quantity $a = \mathcal{N}^m/R^n$ from the observed values of $\mathcal{N}^m$; a should be a constant for the results for a given type of particle

$$a = \mathcal{N}^m/R^n = km^{1-n}. \quad \dots\dots(3)$$

The values obtained are shown in table 5 and in figure 3 for both π - and μ -mesons.

Table 5

Range (microns)	$a_\mu = \frac{\mathcal{N}_\mu^m(R)}{R^{0.711}}$	$a_\pi = \frac{\mathcal{N}_\pi^m(R)}{R^{0.711}}$
50	3.73 ± 0.10	4.10 ± 0.10
100	3.94 ± 0.08	4.39 ± 0.08
150	3.92 ± 0.07	4.55 ± 0.08
200	3.95 ± 0.06	4.57 ± 0.08
250	3.98 ± 0.06	4.58 ± 0.08
300	3.95 ± 0.05	4.58 ± 0.08
350	3.97 ± 0.05	4.50 ± 0.08
400	3.94 ± 0.05	4.52 ± 0.08
450		4.52 ± 0.08
500		4.53 ± 0.08
550		4.55 ± 0.08
600		4.51 ± 0.08
650		4.49 ± 0.08

Weighted means : $a_\mu = 3.95 \pm 0.05$;
 $a_\pi = 4.53 \pm 0.08$.

Table 6

Individual values of the ratio m_π/m_μ

Event No.	m_π/m_μ	Error (%)	Normalized "weight"
2	2.06 ± 0.36	17.4	1.24
14	1.52 ± 0.27	17.6	1.21
8	2.04 ± 0.45	17.7	1.19
12	1.07 ± 0.19	18.0	1.15
5	1.75 ± 0.33	18.7	1.07
17	1.45 ± 0.28	19.3	1.00
8	1.84 ± 0.38	20.5	0.89
9	1.65 ± 0.36	21.8	0.79
1	1.37 ± 0.30	22.0	0.77
4	1.82 ± 0.42	23.3	0.69

Weighted mean : 1.66 ± 0.11 .

In this case, also, the different values are not independent, so that the "probable error" is not reduced by taking the mean.

From equation (3) we can write

$$a_\pi/a_\mu = [m_\pi/m_\mu]^{0.289} = 1.145 \pm 0.024; \quad m_\pi/m_\mu = 1.62 \pm 0.12.$$

We can calculate values of $\mathcal{N}_\pi^m$, corresponding to the observed values of $\mathcal{N}_\mu^m$, from the formula $N = km^{0.289}R^{0.711}$, assuming that $m_\pi/m_\mu = 2.1, 1.65, 1.32$ in turn. The corresponding values of the quantity $\mathcal{N}_\pi^m/R^n$ are represented by the dotted lines in figure 3: (a) corresponds to the case in which the π -meson disintegrates to produce a neutral meson, equal in mass to the μ -meson, while (c) corresponds to the assumption that the momentum balance in the μ -decay is provided by an emitted photon.

§ 6. INDIVIDUAL VALUES OF THE RATIO m_π/m_μ

By using the equation $N = km^{0.289}R^{0.711}$ we can determine a value of m_π/m_μ from the observations on the individual examples of the μ -decay, and the values thus obtained are summarized in table 6. The results are also represented in figure 4. In this diagram the size of the oblong, corresponding to a particular determination, has an area proportional to its statistical weight. It will be seen that of the ten individual determinations, only one has a magnitude less than 1.32, and not one greater than 2.1.

§ 7. THE EXISTENCE OF A NEUTRAL MESON

The three methods which have been employed in the analysis of our observations are consistent in that they show no significant differences from the mean value 1.64 ± 0.11 . Although the fluctuations of the individual values are so large that we cannot regard our results as final, they strongly suggest that the momentum balance in the process of μ -decay is provided by a neutral particle and not by a photon, and that this particle is less massive than the μ -meson. The application of the conservation laws of energy and momentum allows us to calculate the mass, m_ν , of this "neutretto", and we thus obtain the value $m_\nu = 115 \pm 30 m_e$.

Recent photographs by Anderson, Adams, Lloyd and Rau (1947) show that the cosmic-ray mesons brought to rest in the gas of an expansion chamber decay with the emission of an electron of energy 25 mev. Assuming the mass of these mesons to be $200 m_e$, this result is consistent with the assumption that a neutral particle with a mass of about $130 m_e$ is ejected in the opposite direction to the electron in the decay process. The agreement, within the limits of experimental error of the masses of the neutral particles indicated by the two experiments, is suggestive. We have therefore considered the consequences of assuming that the two particles are of the same type, and that the ordinary cosmic-ray mesons are identical with our μ -mesons. The decay sequence that follows from these assumptions is represented in figure 5. Since the electron has half-integral spin we can conclude from the process $\mu \rightarrow e + \nu$ that μ - and ν -mesons differ by half-integral spin. It then follows, from the process $\pi \rightarrow \mu + \nu$, that π -mesons have half-integral spin.

§ 8. THE DECAY CONSTANTS OF MESONS

(a) π -mesons

The distribution in range of the μ -mesons, of which details are given in table 1, are represented in figure 6(a). Following Livingston and Bethe, we can define a straggling coefficient, s , by the relation $s = \sqrt{(\pi/2)} \sqrt{\{\sum_n (R_i - R_m)^2/n\}}$, where R_i is the range of the i th μ -meson, the observed mean value of all the n μ -mesons

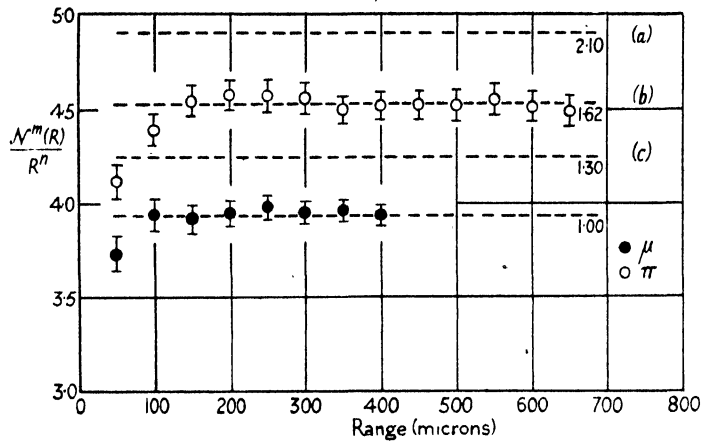


Figure 3. $n=0.711 \pm 0.010$; $\mathcal{N}(R)=km^{1-n}R^n$.



Figure 5.

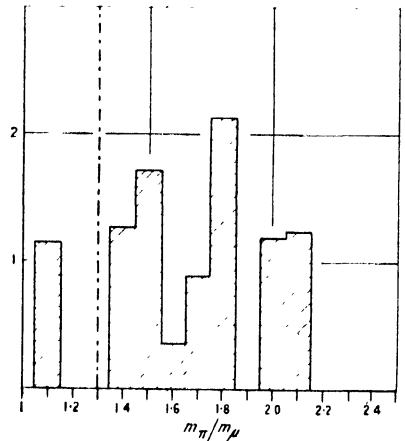


Figure 4. Individual values of mass ratio, m_π/m_μ , as determined by grain-counts.

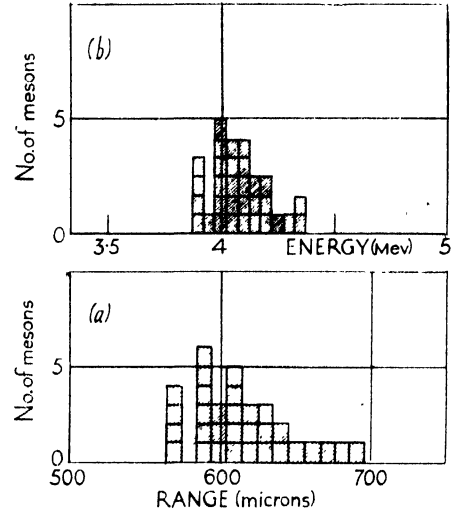


Figure 6. (a) Distribution in range of μ -mesons arising from the decay of π -mesons at rest. (b) Distribution transformed to an energy scale, using extrapolated range-energy relation for protons.

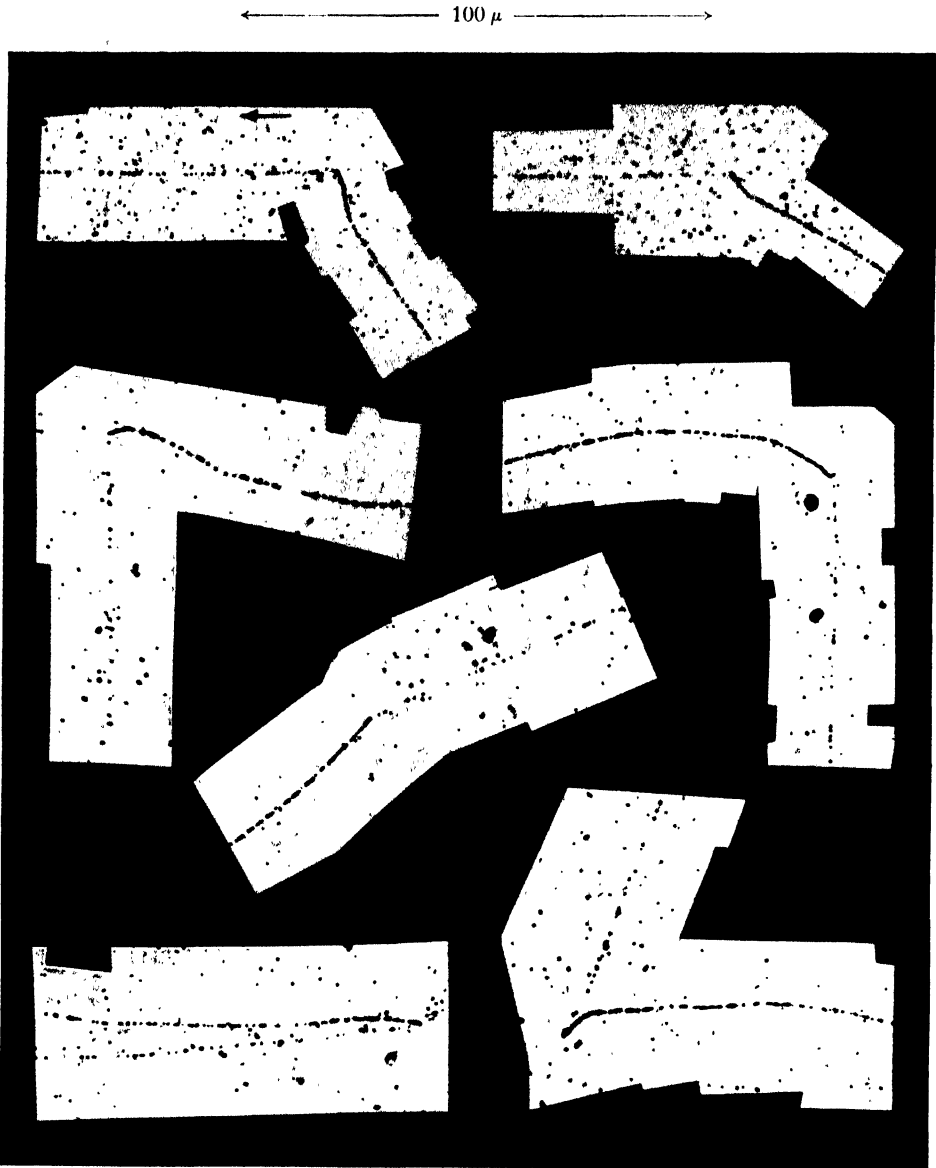


Figure 7. Seven examples of photo-micrographs of the decay of π -mesons, showing that the direction of motion of the μ -meson passes within 1μ of the last grain in the track of the π -meson.

being R_m . The measurements shown in table 1 are thus found to correspond to a value of s given by $s/R_m = 6.24\%$.

Only extrapolated values of the range-energy relation are available, for particles of high velocity and charge $|e|$, in the "Nuclear Research" emulsions. These are given with sufficient accuracy by the equation $E = 0.262 M^{0.425} R^{0.575}$, where E is the energy in mev. and R the range in microns of a particle of M mass-units (proton = 1). It follows from this equation that $\delta E/E = 0.575 \delta R/R$. Writing $\delta R/R = s/R_m = 0.0624$ we find, that $\delta E/E = 0.0359$. This value does not differ significantly from that corresponding to the width at half-maxima of the distribution in energy of the μ -mesons shown in figure 6(b), which has been calculated from the observed range distribution on the assumption that the mass of the μ -mesons equals $200 m_e$. The new observations therefore confirm the previous conclusions that the μ -mesons are emitted with values of the velocity which are constant within narrow limits, for the observed straggling coefficient is of the same order of magnitude as that suggested by the classical approach to the problem.

Now the directions of ejection of the μ -mesons are oriented at random with respect to the line of motion of the μ -mesons at the extreme end of their range. Let the difference in the direction of motion of the two particles, at the point of juncture of the two tracks, be θ . If the π -mesons come to "rest" before decaying, this angle is without significance, for the particles will suffer frequent changes in direction, due to scattering, when moving at low velocities at the end of the range. If, however, a π -meson decays in flight, we should expect the velocity of ejection of the μ -meson, relative to the emulsion, to depend on the value of θ . In these circumstances, the range of the ejected mesons would be very sensitive to the direction of emission relative to the line of motion of the parent π -meson.

We have calculated the magnitude of the effect to be anticipated, making the simplifying assumption that the π -mesons always decay when 5μ from the end of their range. It is easily shown that in these circumstances the distribution in range of the μ -mesons would extend from about 200μ to 1200μ . Further, we have examined the distribution in range of the μ -mesons in these events for which $\theta < \pi/2$, and, separately, those for which $\theta > \pi/2$. It is found that there is no significant difference between the two distributions. We therefore conclude that the π -mesons, in the observed events, were at or very near the end of their range when they decayed.

We cannot conclude, however, that the observed events represent all the π -mesons which, having decayed in the emulsion, led to the production of a μ -meson which stopped in the emulsion. If, for example, a π -meson decayed when its residual kinetic energy exceeded 2 mev., it would be difficult to distinguish the event from one in which a meson is scattered by a nuclear collision. We believe, however, that if the residual range of the π -meson were less than 100μ at the instant of decay, we should recognize the event because of the sudden decrease of the grain density in the track immediately after the point of "scattering". We therefore conclude that in twenty cases a π -meson traversed a distance of 100μ at the end of its range without decaying, and that in no case did it decay in this interval. The time of flight of such a meson is 3×10^{-12} sec., and we can therefore write $20 \times 3 \times 10^{-12} \times k_\pi < 1$, where k_π is the decay constant of the π -mesons. It follows that $k_\pi < 1.6 \times 10^{10}$ and $\tau_\pi > 0.4 \times 10^{-10}$ sec.

In every case in which the μ -meson reaches the end of its range in the emulsion it stops without leading to a disintegration with the emission of heavy particles. "This suggests that the π - and μ -particles are positively charged—for the Coulomb repulsion would then prevent their close approach to a nucleus when moving with low velocities, and their interaction with it—and we shall assume this to be the case.*

We have now observed 150 events in which a π -meson, at the end of its range, produces a μ -meson. In every case we can distinguish no displacement of the line of motion of the ejected μ -meson from the last grain in the track of the π -meson (see figure 7). We regard this as evidence that in no case did the π -meson, at the end of its range, diffuse a distance greater than 2μ before decaying. If it is assumed that a positively charged positive meson diffuses through the solid material of the emulsions as through a heavy gas, an estimate can be made of an upper limit to the half-period value τ_π . The result so obtained is $\tau_\pi > 5 \times 10^{-7}$ sec. Dr. Frohlich has pointed out to us, however, that the problem of the motion of a positively charged meson in a solid substance is one of great complexity. It is likely that a positive meson will be "trapped" when brought to rest in a solid and that its diffusion will thus be prevented. We can therefore attach little weight to the limit obtained by this method.

(b) μ -mesons

We have seen that in thirty cases the μ -mesons reach the end of their range in the emulsion. In no case have we observed the track of a secondary μ -meson to terminate in the emulsion without having the characteristics associated with that of a particle at the end of its range. We can therefore write for the decay constant k_μ of these mesons $30k_\mu \times 10^{-11} < 1$, i.e. $k_\mu < 3.3 \times 10^9$; $\tau_\mu > 2 \times 10^{-10}$ sec. The lower limits to the values of τ_π and τ_μ derived from these observations have the following important consequence. They indicate that a meson, of either type, with an energy of the order of 4 mev., will be unlikely to decay in free flight, *in vacuo* or in a gas at normal pressures over a length of path p of at least several millimetres. Thus for μ -mesons we have $p_\mu > 8.7 \times 10^9 \times 2 \times 10^{-10}$ cm. = 1.7 cm., and for π -mesons $p_\pi > 6.7 \times 10^9 \times 0.6 \times 10^{-10}$ cm. = 4 mm. It is therefore reasonable to undertake magnetic deflection experiments which demand for their success that the particles should not suffer spontaneous decay in a length of path of the order of 1 or 2 cm. (Powell and Rosenblum 1948).

(c) *The postulated neutral meson*

The observed degree of homogeneity of the range distribution of the μ -mesons produced by the decay of π -mesons at rest indicates that the kinetic energy of the ejected particles E_μ is subject to a variation of less than 0.15 mev., or 2.3×10^{-7} ergs. Further, the lower limits to the lifetime of the π - and μ -mesons show that the energies to which their rest-masses are equivalent must be constant to within

* These observations do not, of course, provide conclusive evidence for the sign of the charge, for the μ -mesons may have a very weak interaction with nuclei. Further, although evidence is accumulating in support of our original suggestion that they are identical in type with the mesons commonly observed in experiments with expansion chambers and counters, there is no decisive evidence available to show that negatively charged mesons of this type, when brought to rest in materials of high atomic number, produce disintegrations with the emission of heavy particles. All that is known definitely, at the present time, is that they disappear without giving rise to a delayed coincidence.

10^{-17} ergs. Any variation, δE_μ , in the energy of ejection of the μ -meson, of the order of 10^{-7} ergs, would be associated; therefore, with a variation in the mass of the postulated neutral meson. It follows that the observed degree of constancy of the values of E_μ allows us to set a lower limit to the half-value period of the neutral meson τ_ν . We write $\delta E \delta t \sim h$, where h is Planck's constant. Since $\delta E \leq 2.3 \times 10^{-7}$ ergs, $\tau_\nu \geq 2.6 \times 10^{-21}$ sec.

REFERENCES

- ANDERSON, C. D., ADAMS, R. V., LLOYD, P. E., and RAU, R. R., 1947, *Phys. Rev.*, **72**, 724.
 LATTES, C. M. G., OCCHIALINI, G. P. S., and POWELL, C. F., 1947, *Nature, Lond.*, **160**, 453, 486.
 POWELL, C. F., and ROSENBLUM, S., 1948, *Nature, Lond.*, **161**, 473.

Determination of the Masses of Charged Particles Observed in the Photographic Plate

By Y. GOLDSCHMIDT-CLERMONT*, D. T. KING,
H. MUIRHEAD AND D. M. RITSON†

H. H. Wills Physical Laboratory, Bristol University

* On leave of absence from Brussels University

† On leave of absence from Clarendon Laboratory, Oxford

Communicated by C. F. Powell; MS. received 10 April 1948

ABSTRACT. A study has been made of scattering of 40 protons and 160 mesons in photographic emulsions in order to determine the mass of the individual particles. The spread in the values so obtained is large, but the evidence suggests that the majority of mesons recorded by the emulsion can be identified both with the mesons, of mass $\simeq 200 m_e$, commonly observed in cloud-chamber experiments and with counters, and with the μ -mesons observed in the photographic plates. The slow particles producing nuclear disintegrations, σ -mesons, appear to contain a large proportion of particles with a mass equal to that of the π -mesons.

§ 1. INTRODUCTION

PREVIOUS work (Bose and Chowdhry 1941, Perkins 1947, Occhialini and Powell 1947) has shown that some of the characteristics of the particles observed in photographic emulsions can be determined by an inspection of the "grain-density" and multiple Coulomb scattering associated with the tracks caused by the passage of these charged particles. Such methods can be extended to give quantitative information about the masses and charges of the particles.

Grain counting is especially favourable for the determination of the ratio of the masses of particles involved in contemporaneous events, such as the π - μ meson decay process (Lattes, Occhialini and Powell 1947); but it is not well suited for absolute mass determinations, for reasons discussed by the authors.

The method based on observations of multiple Coulomb scattering is independent of the fading of the latent image and gives absolute values. Although inherently less accurate than grain counting, it is therefore more suited for work with particles occurring in events of which the relationship in time is unknown.

The theory of multiple scattering has been developed by Williams (1939, 1940), and several experiments have been made to verify the formulae which he established. Probably the most accurate and complete is that of Kulchitsky and Latyshev (1942) on the scattering of fast electrons by light and heavy elements. These authors found agreement with Williams' formula for light elements, but noted systematic deviations for the heavier elements. These deviations can be explained by the more accurate calculations of Molière (1947) on the small angle single scattering by a screened Thomas-Fermi atom.

It would appear, therefore, that the calculations of Williams and of Molière form a reliable basis for determining masses by observations on multiple scattering.

§ 2. THEORY OF METHOD

A charged particle traversing a medium suffers frequent small-angle deflections in elastic collisions with the nuclei of the atoms composing the medium. As a result of large numbers of these small deflections, the track has a curved appearance. This scattering is purely statistical in nature, and Williams has evaluated the probability of a change of direction, α , occurring in a given medium, for a path length t , from the distribution of the values of the angles of scattering occurring in single collisions. Williams' theory shows that the probability distribution for small values of α is very nearly Gaussian; for large values the distribution approaches that given by the Rutherford formula for single scattering. The deflections which he considers are the projection of the true deflections on a two-dimensional plane. In our experiments, also, we define α as the two-dimensional projection of the true deflection.

The arithmetic mean value of the deviations considered, $\langle \alpha \rangle_{AV}$, is given by a formula of the form $\langle \alpha \rangle_{AV} = \{f(\beta, t, N, Z)\}/E$, where β = ratio of velocity of particle to velocity of light, t = thickness of scattering medium traversed by particle, N = number of atoms per cm^3 of the scattering medium, Z = charge number on nucleus of atom of scattering medium, E = mean kinetic energy of particle in element of trajectory under consideration.

α varies with the energy and accordingly with the residual range of the particle. It is convenient to define a quantity, ϵ , which is independent of range and which is given by the relation $\epsilon = \alpha E / f(\beta, t, N, Z)$. The values of ϵ will have a probability distribution similar to that for α , and $\langle \epsilon \rangle_{AV} = 1$.

Consider the trajectory of a charged particle, of unknown mass, moving in a given medium and undergoing successive deflections $\alpha_1, \alpha_2, \alpha_3$ etc. If we assume the particle to have a mass M , we can, from the range-energy curve for protons, calculate the velocity of the particle for any value of the residual range. It is then possible to calculate the quantity $\langle \epsilon \rangle_{AV}$ for the trajectories of particles of this type; and similar calculations can be made for any assumed value of the mass. The value of M , which enables us to satisfy the relation $\langle \epsilon \rangle_{AV} = 1$, for the trajectory of a particular particle, then gives a measure of its mass.

M can be determined analytically in the following manner. We evaluate $\langle \epsilon_m \rangle_{AV}$ for an approximately correct value m of the mass.

Now

$$\epsilon_M = \epsilon_m \frac{E_M}{E_m} \cdot \frac{f(\beta_m, t, N, Z)}{f(\beta_M, t, N, Z)}.$$

$f(\beta, t, N, Z)$ is relatively insensitive to a change in mass, in comparison with E . We can therefore write $\epsilon_M \simeq \epsilon_m \cdot E_M / E_m$. The relationship between mass

and energy may be calculated from the range-energy data for photographic emulsions (Camerini 1947). This gives $E_M/E_m \simeq (M/m)^{0.44}$, since the range exponent is 0.56 for particles of charge $|e|$. Therefore $\epsilon_M \simeq (M/m)^{0.44} \epsilon_m$ and $\langle \epsilon_M \rangle_{AV} = (M/m)^{0.44} \langle \epsilon_m \rangle_{AV}$. But by definition $\langle \epsilon_M \rangle_{AV} = 1$, so that $M = m \{ \langle \epsilon_m \rangle_{AV} \}^{-2.23}$.

The values of ϵ are distributed according to a probability function and the mean, $\langle \epsilon \rangle_{AV}$, is a statistical quantity. It is therefore important to decide whether its value is best determined from the arithmetic mean, mean square or median of the individual values. An analysis shows that the arithmetic mean of all values of ϵ below a certain cut-off point should be determined in order to give the least probable error. This mean, which we shall denote as $\langle \epsilon_{CO} \rangle_{AV}$, is the arithmetic mean of all values of ϵ less than $4 \langle \epsilon_{CO} \rangle_{AV}$. The determination of $\langle \epsilon_{CO} \rangle_{AV}$ depends upon an approximate knowledge of the answer; in practice the problem is rapidly solved by a relaxation method. The necessity for ignoring terms greater than $4 \langle \epsilon_{CO} \rangle_{AV}$ arises from the fact that very large deviations due to single collisions may occur with finite probability; and these would, if included in the statistics, give rise to very large fluctuations in mass values.

The distribution of the individual values of ϵ is very nearly Gaussian. From an examination of this distribution, for ϵ up to the cut-off point, the probable statistical error in the determination of $\langle \epsilon_{CO} \rangle_{AV}$ was found to be $\pm (0.55/\sqrt{n}) \langle \epsilon_{CO} \rangle_{AV}$, where n is the number of independent observations made on the deflections. When n has large values, this leads to a probable error in the mass determinations of $\pm 123/\sqrt{n} \%$.

§ 3. EXPERIMENTAL

After attempts had been made to measure the scattering along a track in a photographic emulsion by means of a normal high magnification microscope, it was decided to use the mechanized projection microscope (Occhialini, Payne and Powell 1948) for this purpose. Large-scale drawings were taken with this instrument. The advantage thus gained over direct measurements were threefold: firstly, the spurious scattering on the track was reduced (see below); secondly, the residual spurious scattering could be accurately evaluated; and lastly, a permanent facsimile of the track at high magnification was available for inspection and analysis.

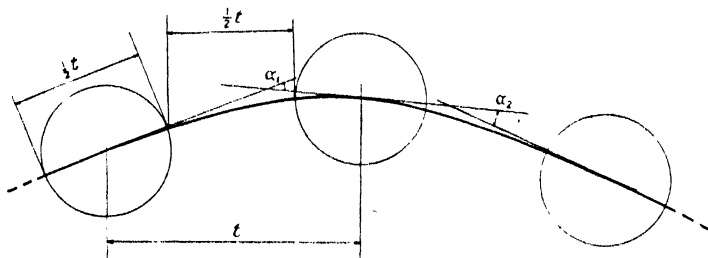
In order to facilitate the actual drawing, a magnification of $\times 4000$ was used. This magnification was obviously too large relative to the resolving power, but it allowed us to obtain drawings on a convenient scale for analysis. Only the centre of the field of the projection microscope was used in making the drawings.

The drawing of the track was then divided up into a series of segments, the lengths of which were chosen to fulfil the conditions indicated below. The changes in direction between the middle points of alternate segments were then measured. The distances between these middle points are the cell lengths to which reference is made below.

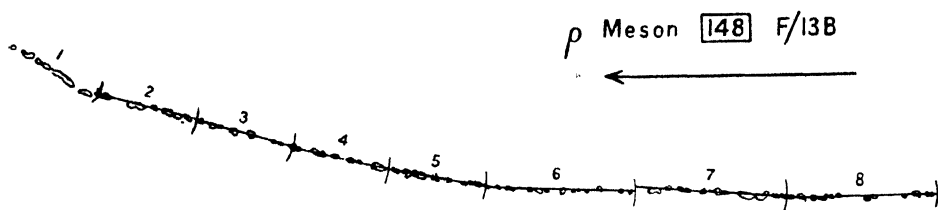
There were two possible methods of measurement: of angles either between successive tangents, or between successive chords ("sagitta" measurements are a special case of chords). Since the track is defined by a small number of grains, it is obviously impossible to draw accurate tangents. The most accurate procedure is to introduce deliberately a small systematic error into the measurements by measuring angles between mean lines drawn through grains over certain segments of a track. The length over which this mean line is drawn is chosen as half the

cell length (figure 1). The mean line will generally be intermediate between a chord and tangent. The theory of a multiply scattered track has been given by Rossi (1941) and leads to an easy calculation of the systematic error, as shown in Appendix 1. It is found that measured angles must be increased by $4 \pm 2\%$ to correspond with those in Williams' theory (Appendix 2).

We have found it possible to increase the effective number of measurements on a single track by the following procedure, illustrated in figure 2. The track is



A. Method of measurement of angular deviations.



B. Reproduction of section of drawing made on typical meson track, showing lines drawn through grains in successive segments.

Figure 1.

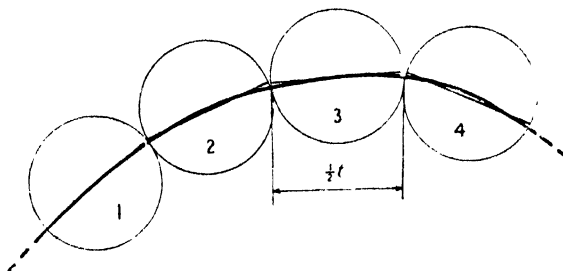


Figure 2. Overlapping half cells drawn on track.

divided into overlapping half cells, 1, 2, 3, etc. In each of these a straight line is drawn to pass, as closely as possible, through the centres of gravity of the grains. The angular deviations between the lines in 1 and 3, 2 and 4, 3 and 5, etc., are then measured. It can be shown that for a large number of observations of this type two-thirds of the total number of measurements are statistically independent.

The cell length used must be such as to keep the experimental errors low compared with the expected scattering. The experimental errors arise from the finite size of the grains, the distortion of the microscopic field, the errors in drawing and the errors in measuring. These effects give a "spurious scattering" superimposed on the Coulomb multiple scattering.

The spurious scattering has been evaluated by means of observations on the tracks of very fast particles (protons of 40 mev., one α -particle of 120 mev.). These tracks have a negligible multiple scattering, and, in the case of protons, the same grain density at corresponding points in the tracks as mesons of the same velocity. Almost all the changes in direction observed in measurements on these tracks are due to the sources of error indicated above; they correspond to spurious scattering. The results of these measurements are given in Appendix 3. They show, as would be expected, that the spurious scattering decreases with longer cells and high grain density. On the contrary, the Coulomb multiple scattering increases with longer cells. The length of the successive cells can therefore be chosen to keep the spurious scattering less than one-quarter of the true multiple Coulomb scattering.

The values chosen are:

Protons :	Number of half cell	1	2	3	4	5	6	7	8	9			
	Length of half cell (μ)	12.5	12.5	12.5	12.5	12.5	12.5	20	20	20			
	Lengths of successive overlapping cells (μ)	—	25	25	25	28.25	36.25	40	40	40			
Mesons :	Number of half cell	1	2	3	4	5	6	7	8	9	10	11	12
	Length of half cell (μ)	10	10	10	10	10	15	15	15	15	15	15	20
	Lengths of successive overlapping cells (μ)	—	20	20	22.5	27.5	30	30	30	30	32.5	37.5	40

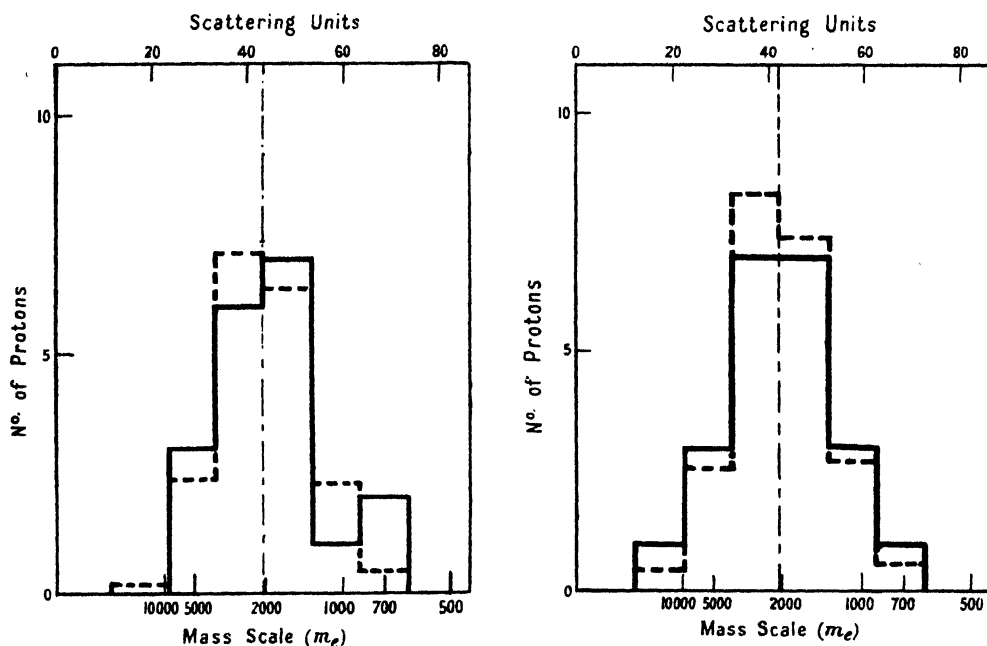
§ 4. RESULTS

As a check on the method, measurements were made on two series of protons. Twenty proton tracks (each 300 microns in length) were selected at random from a plate exposed to the disintegration particles arising from the reaction $B^{10}(d, p)B^{11}$. The mean value obtained for the experimental mass was 1.10 ± 0.10 proton masses. This value depends upon a knowledge of the atomic composition of the emulsion, and we have used the values given by Messrs. Ilford Ltd. The limits of error indicated correspond to statistical fluctuations alone. There is an additional uncertainty, due to the inherent errors of the experiment, equal to about ± 0.07 proton masses.

Similar observations on twenty proton tracks observed in cosmic-ray plates gave a value 1.20 ± 0.10 proton masses. There is no significant deviation from the expected value, particularly as we cannot exclude the possibility that tracks due to deuterons or tritons may have been included among those selected for measurement. Histograms of the distribution of the individual values of the mass are shown in figure 3 by full lines. The dotted lines show the corresponding distribution calculated for a gamma distribution and normalized to correspond to the same number of tracks. A comparison of the two distributions indicates a satisfactory internal consistency between the observations. The mean $\langle \epsilon_{CO} \rangle_{AV}$ was cross-checked for constancy with range and varying cell size by taking the values of ϵ at the same residual range for all the particles of the same class, and then determining $\langle \epsilon_{CO} \rangle_{AV}$ for these terms. The values of the means thus obtained show no significant variations.

The tracks of particles on the cosmic-ray plates (Ilford, Boron-loaded C.2 exposed at 2800 m. and 5500 m.) with grain density less than those of protons were divided into four groups. The π^- and μ -mesons were taken from the π - μ decay process. Some π -mesons were taken in which the μ -meson resulting from

the decay did not end in the emulsion. In these cases, however, only those π -mesons were accepted for measurement of which the track of the associated μ -meson was of sufficient length to allow a definite identification. Similarly, a μ -meson track was accepted for measurement only if it was accompanied by the track of the primary π -meson, and if the track of the μ -meson of length of the order of 600 microns terminated in the emulsion. Following the previous definition adopted in this laboratory (Lattes, Occhialini and Powell 1947), we distinguish a



A. 20 protons from H.T. set.

B. 20 protons from cosmic rays.

Figure 3. Histograms of the distribution of values for proton masses, in scattering units and mass units. The broken line represents the expected distribution; full line represents the observed distribution.

third class of mesons, σ -mesons, which suffer capture and lead to observable nuclear disintegrations in the emulsion. We can also define a fourth group, ρ -mesons, which stop in the emulsions without giving rise to any observed charged particles. The origin of these mesons is uncertain, but most of them can probably be identified as the mesons commonly observed in experiments with Wilson chambers and counters.

Combination of the scattering results with previous grain-density measurements permits an unequivocal assignment of charge $|e|$ to all measured particles. If the charge were greater, the results would lead to a gross inconsistency between the masses as measured by scattering and by grain counting.

The results of the mass determinations on 20 π -, 20 μ -mesons, 60 σ - and 60 ρ -mesons are shown below together with the "probable" statistical error.

Mass of π -mesons: $260 \pm 30 m_e$,

Mass of μ -mesons: $205 \pm 20 m_e$,

Mass of σ -mesons: $275 \pm 15 m_e$,

Mass of ρ -mesons: $200 \pm 10 m_e$.

The systematic experimental error is again of the order of $\pm 7\%$. Histograms similar to those for the protons are given for these four classes in figure 4.

The histogram for the ρ -mesons shows a slight tailing at the low-mass end, two mesons lying outside the expected histogram. The statistical errors are, however, too great to allow us to regard these observations as evidence for the existence of

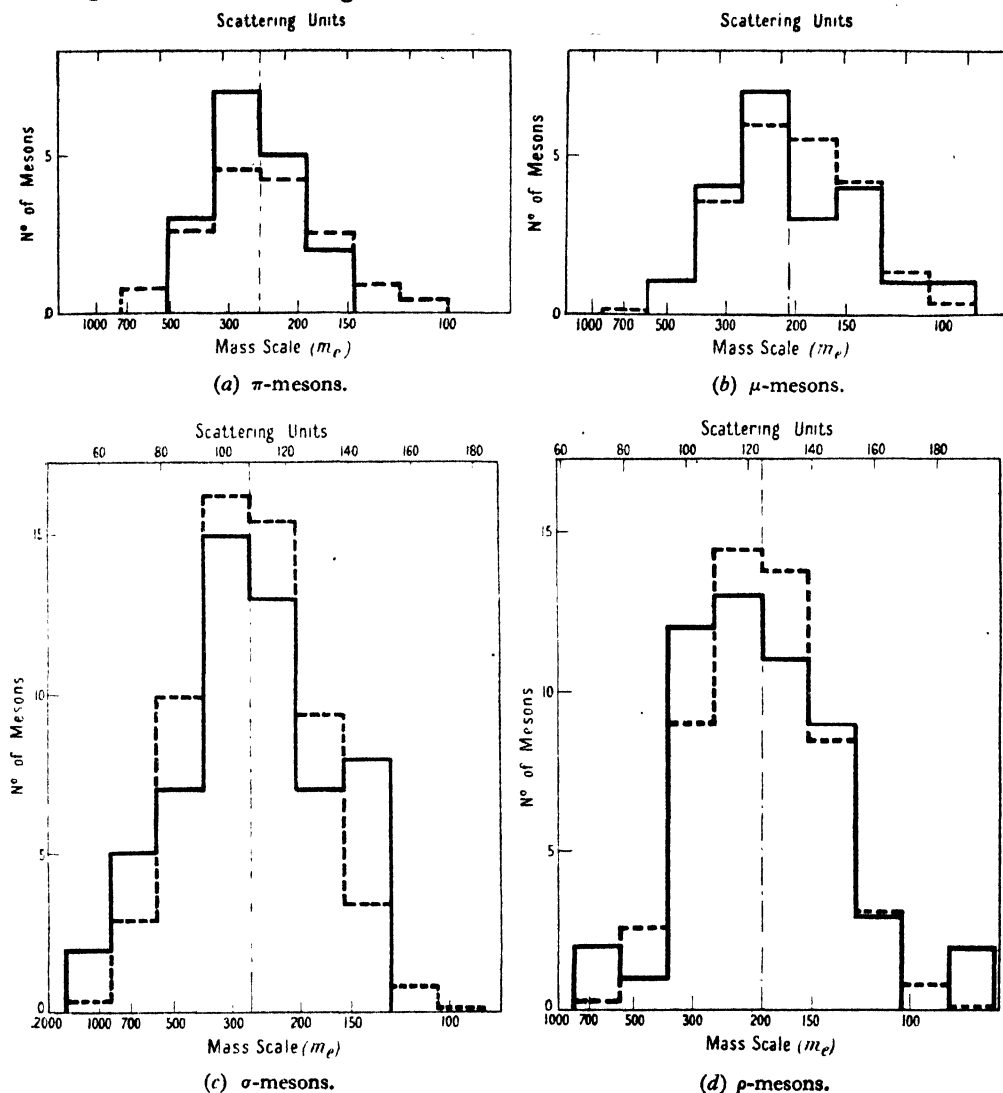


Figure 4. Expected (broken line) and observed (full line) histograms for mesons.

mesons of mass approximately $80 m_e$. The distribution in ϵ is not "normal" as the method of taking overlapping cells gives rise to fluctuations in the number of independent readings, and the imposition of a "cut-off" will also give rise to occasional large values of $\langle \epsilon_{CO} \rangle_{AV}$. The theoretically expected histogram was calculated by taking the normal frequency distribution with the same probable error. In practice for the above-mentioned readings a slight tailing of such a distribution should occur. The results otherwise appear consistent with the expected probable error.

§ 5. CONCLUSION

It has previously been suggested (Lattes, Occhialini and Powell 1947, Marshak and Bethe 1947) that the constitution of the π -, μ -, σ - and ρ -mesons is as follows :

The π -mesons are heavy mesons with masses of 300–420 m_e and strong nuclear interaction. The μ -mesons are identical with those hitherto observed in cloud chamber and delayed coincidence experiments with approximate mass 200 m_e (Fretter 1946, Hughes 1947, Valley 1947). The ρ -mesons were supposed to be predominantly μ -mesons with a slight admixture of π -mesons for which the decay had escaped observation, and the σ -mesons to contain a high proportion of negative π -mesons.

Our results give support to these assumptions. The mean mass of the σ -mesons is significantly higher than that of the ρ -mesons. The masses of the μ - and ρ -mesons are in agreement, within the probable error, with the results of Fretter's mass determinations at sea level (mass $202 \pm 5 m_e$). The best value of the mass ratio of π - to μ -mesons from grain counting is 1.66 ± 0.11 (Lattes, Occhialini and Powell 1948). There is no significant disagreement with the mass ratio obtained from our experiments of 1.3 ± 0.2 .

APPENDIX 1

Correction for systematic errors

The number of useful independent measurements which can be made on the track of a particle is controlled by two conflicting factors. To minimize the statistical errors, it is desirable to make the largest possible number of independent observations on the scattering, and therefore to choose the smallest possible cell length. In practice, however, the effects of spurious scattering impose a lower limit to the length of the cell which can be employed in a given region of the track. The error introduced in the evaluation of $\langle \epsilon_{CO} \rangle_{AV}$ is the result of a compromise between the errors due to statistical fluctuations and to measurement.

In conformity with these considerations, cell lengths along the track are chosen so that at each point the spurious scattering does not exceed a known small fraction of the multiple scattering. Within each cell the deviation is measured with the least possible error by drawing the mean line through the grains contained in successive segments of the track and measuring the angular change between the directions of the track in these segments. A limit is set to the size of the segment by the magnitude of the systematic errors introduced. These systematic errors arise in the following way. It can be proved quite generally that the arithmetic mean of the angular deviations between successive tangents along the multiply scattered track is larger than the arithmetic mean of the deviations between successive chords. The mean line drawn through consecutive grains will in some cases correspond more closely to a tangent and in other cases to a chord. The arithmetic mean of such a series of measurements, $\langle \alpha_{exp} \rangle_{AV}$, will have upper and lower limits set by the values for the means of tangents and chords. The calculation of the limits is most easily carried out by the methods given by Rossi and Greisen (1941).

Consider figure 5. Chords AB and CD of length kt and tangents A'B and CD' are drawn. The cell length is t ; kt is a fraction of this length. We require to

find the mean value $\langle \alpha \rangle_{AV}$ of the angle $\overline{BA} - \overline{CD}$. Now

$$\alpha_{exp} = \overline{BA} - \overline{CD} = (\overline{BA} - \overline{B'A}) + (\overline{B'A} - \overline{CD'}) - (\overline{CD} - \overline{CD'}),$$

and, therefore,

$$\langle \alpha_{exp}^2 \rangle_{AV} = \langle (\overline{BA} - \overline{B'A})^2 \rangle_{AV} + \langle (\overline{B'A} - \overline{CD'})^2 \rangle_{AV} + \langle (\overline{CD} - \overline{CD'})^2 \rangle_{AV} \dots (1)$$

The cross-terms vanish in the average as they contain positive and negative contributions of equal magnitude. Rossi and Greisen (1941) showed that

$$\langle (\overline{CD} - \overline{CD'})^2 \rangle_{AV} = \langle (\overline{BA} - \overline{B'A})^2 \rangle_{AV} = \frac{1}{3} \alpha_0^2 kt,$$

where α_0^2 is the mean square of angles between successive tangents at intervals of one unit of length. Similarly $\langle (\overline{B'A} - \overline{CD'})^2 \rangle_{AV} = \alpha_0^2 t(1-k)$. Substituting these values in equation (1), $\langle \alpha^2 \rangle_{AV} = \alpha_0^2 t(1 - \frac{1}{3}k)$.

As was explained above, $\langle \alpha_{exp} \rangle_{AV}$ has upper and lower limits corresponding to the mean deviations between successive chords and tangents. The mean square deviation for the tangent is by definition $\alpha_0^2 t$, and, therefore,

$$\sqrt{(2/\pi)\alpha_0\sqrt{t}} < \langle \alpha_{exp} \rangle_{AV} < \sqrt{(2/\pi)\alpha_0\sqrt{t}(1 - \frac{1}{3}k)},$$

since the transformation from mean square to arithmetic mean involves a constant which is the same throughout the terms.

The most suitable value for k from the experimental data available was taken as $k = \frac{1}{3}$. For this value $\sqrt{(2/\pi)\alpha_0\sqrt{t}} < \langle \alpha_{exp} \rangle_{AV} < \sqrt{(2/\pi)0.92\alpha_0\sqrt{t}}$, and, therefore,

$$\langle \alpha_{exp} \rangle_{AV} = 0.96\sqrt{(2/\pi)\alpha_0\sqrt{t}} \pm 0.02\sqrt{(2/\pi)\alpha_0\sqrt{t}}.$$

Similar calculations were made for sagitta with smoothing as shown in figure 6. This treatment leads to difficulties in calculation of the scattering distribution



Figure 6. Sagitta measurement on smoothed track. Vertical line indicates sagitta distance.

function; the connection between tangents and sagitta was only derived by Rossi by effectively ignoring the single scattering "tail". At the same time, the accuracy of the measurement is not improved. The method described above was therefore adopted in preference to sagitta.

APPENDIX 2

Application of Williams' theory to mixed media

The extension of Williams' theory to media consisting of a mixture of atoms is comparatively simple. Williams gives the following summary of his method. The units used are δ units where, $\delta = 2Ze^2(Nt)^{1/2}/(M\beta^2c^2\xi)$; Z is the atomic number of scattering nuclei; ze , M and βc are charge, mass and velocity respectively of

the scattered particle, $\xi = (1 - \beta^2)^{1/2}$; N is number of atoms per cm^3 ; and t is thickness of scattering material.

The arithmetic mean scattering in units of δ is given by Williams as

$$\langle \alpha \rangle_{AV} = 0.80 \langle \alpha_m \rangle + 1.45,$$

$$\langle \alpha_m \rangle^2 = 2 \ln [\phi_{\max} / \phi_{\min}].$$

Williams defines $\phi_{\max}$ as equal to $\sqrt{(\pi/2)} \delta$ -units and $\phi_{\min}$ as a characteristic angle for the atoms of the media, given (by Williams) for low and high velocities of the incident particles. Molière (1947) has calculated $\phi_{\min}$ for all velocities.

Before Williams' theory can be applied to mixtures, the following changes must be made, δ being defined as

$$\delta = \left(\sum_i \frac{4Z_i^2 x^2 e^4 N_i t}{M^2 \beta^4 c^4 \xi^2} \right)^{1/2},$$

where the subscript i refers to the different molecular species. $\langle \alpha_m \rangle^2$ now becomes

$$\frac{8x^2 e^4 t}{M^2 \beta^4 c^4 \xi^2} \cdot \frac{\sum_i N_i Z_i^2 \ln [\phi_{\max} / (\phi_{\min})_i]}{\delta^2}.$$

As before, $\phi_{\max} = \sqrt{(\pi/2)} \delta$ -units and $(\phi_{\min})_i$ refers to the characteristic angle for each molecular species i in the medium.

For the purpose of finding $\langle \alpha_{CO} \rangle_{AV}$ with a large angle cut-off, the distribution in α is well represented for large values of α by

$$(\pi/\alpha^3)(1 + 3\pi \langle \alpha_m \rangle / \alpha^2).$$

In figure 7, curve A gives the value of $\langle \alpha \rangle_{AV}$, and curve B $\langle \alpha_{CO} \rangle_{AV}$, both measured in terms of δ -units, as a function of cell length in the Ilford C.2 Boron-loaded emulsion. In curve B all angles greater than 12δ are cut off. The value for δ was taken as $1.56\sqrt{t/E}$, where t is in units of 10 microns.

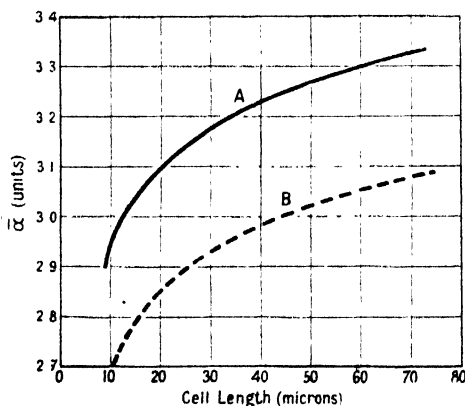


Figure 7. The variation for different cell lengths of (A) the arithmetic mean scattering $\langle \alpha \rangle_{AV}$ in δ -units, (B) $\langle \alpha_{CO} \rangle_{AV}$ in similar units with angles greater than 12δ cut-off.

APPENDIX 3

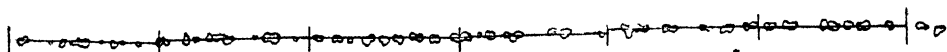
Determination of spurious scattering

The main limitations on the accuracy with which the scattering of a track can be measured are due (a) to the finite size of the grains making up a track, and (b) to the introduction of experimental inaccuracies. These effects produce spurious scattering, whose distribution is approximately a Gaussian error function. The real multiple scattering of the track also has an approximately Gaussian distribution. The experimentally measured scattering is therefore formed by the Gaussian addition of the real and spurious scattering $\langle \alpha^2 \rangle_{\text{exp}} = \langle \alpha^2 \rangle_{\text{real}} + \langle \alpha^2 \rangle_{\text{spur}}$, and, therefore, the quantity required is $\langle \alpha^2 \rangle_{\text{real}} = \langle \alpha^2 \rangle_{\text{exp}} - \langle \alpha^2 \rangle_{\text{spur}}$, which may be written as $\langle \alpha \rangle_{\text{real}} \simeq \langle \alpha \rangle_{\text{exp}} (1 - \frac{1}{2} \langle \alpha^2 \rangle_{\text{spur}} / \langle \alpha^2 \rangle_{\text{exp}})$, for $\langle \alpha \rangle_{\text{exp}} \gg \langle \alpha \rangle_{\text{spur}}$.

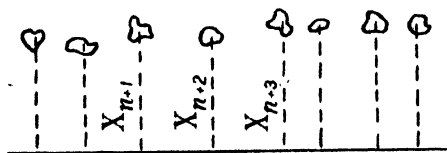
In order to determine $\langle \alpha \rangle_{\text{spur}}$ the following series of experiments was carried out.

Drawings were made, with the aid of the projection microscope, of the tracks of 12 fast protons (*c.* 40 Mev.) and one fast α -particle (*c.* 120 Mev.). Sections of tracks selected for measurement had grain densities varying from 3 to 10 grains per 10 microns, and were traced out under a wide range of optical conditions and with wide variations of the density of the background grains. For these tracks $\langle \alpha \rangle_{\text{spur}}$ was much greater than $\langle \alpha \rangle_{\text{real}}$ and the measurements were effectively a determination of the spurious scattering.

To determine spurious scattering two types of measurements were made. Those shown in figure 8(a) are the same as those previously described in the



(a) Mean lines through grains of overlapping cells for fast particle.



(b) "Triple-point" determinations.

Figure 8. Determination of spurious scattering.

determination of the multiple scattering along the track. Those shown in figure 8(b) consisted of so-called triple-point determinations in which the displacements, X , of the centres of gravity of the grains were measured, from a line running roughly parallel to the direction of the track, and

$$\langle X \rangle_{\text{AV}} = (1/(n-2)) \sum_3^n |(X_{n-2} - X_{n-1}) - (X_{n-1} - X_n)|$$

was measured. $\langle X \rangle_{\text{AV}}$ is, therefore, a factor which indicates the mean displacement of the individual grains about the trajectory of the particle, and provides a useful indication of the spurious scattering on all types of track, even those with large multiple scattering.

The results of the measurements show the following relationship between $\langle X \rangle_{\text{AV}}$ in microns for a grain density of N per 10 microns, cell-length t in 10 microns units (as in figure 1), and $\langle \alpha \rangle_{\text{spur}} = 39 \langle X \rangle_{\text{AV}} / \sqrt{N} \cdot t^{1/2}$, measured in degrees. $\langle X \rangle_{\text{AV}}$ for individual tracks had values ranging from 0.125 micron to 0.25 micron. Measurements on 10 meson tracks showed the same limits. The cell length for measurements on mesons and protons were chosen so that for $\langle X \rangle_{\text{AV}} = 0.25$ micron, $\langle \alpha \rangle_{\text{spur}} / \langle \alpha \rangle_{\text{real}} < \frac{1}{4}$.

The correction for spurious scattering over a whole track was thus maintained at less than 3%.

ACKNOWLEDGMENTS

Our thanks are due to Professor Tyndall for generously extending the facilities of the Bristol Laboratory to two of us (Y.G. and D.M.R.), and to Dr. C. F. Powell and Dr. G. P. S. Occhialini for constant help and encouragement during the course of the work. We are also indebted to Dr. Occhialini for the suggestion of

using the mechanized projection microscope for scattering measurements, and to the work of the observers of Bristol University.

REFERENCES

- BOSE, D. M., and CHOWDHRY, B., 1941, *Nature, Lond.*, **147**, 240.
 CAMERINI, U., and LATTES, C. M. G., 1947, data supplied with Ilford "Nuclear Research" emulsions.
 FRETTER, W. B., 1946, *Phys. Rev.*, **70**, 625.
 HUGHES, D. J., 1947, *Phys. Rev.*, **71**, 387.
 KULCHITSKY, A., and LATYSHEV, G. D., 1942, *Phys. Rev.*, **61**, 254.
 LATTES, C. M. G., FOWLER, P. H., and CUER, P., 1947, *Proc. Phys. Soc.*, **59**, 883.
 LATTES, C. M. G., MUIRHEAD, H., OCCHIALINI, G. P. S., and POWELL, C. F., 1947, *Nature, Lond.*, **159**, 694.
 LATTES, C. M. G., OCCHIALINI, G. P. S., and POWELL, C. F., 1947, *Nature, Lond.*, **160**, 453, 486; 1948 (unpublished).
 MARSHAK, R. E., and BETHE, H. A., 1947, *Phys. Rev.*, **72**, 506.
 MOLIÈRE, G., 1947, *Z. für Naturforschung*, **2 A**, 133.
 OCCHIALINI, G. P. S., and POWELL, C. F., 1947, *Nature, Lond.*, **159**, 186.
 OCCHIALINI, G. P. S., PAYNE, R. M., and POWELL, C. F., 1948 (unpublished).
 PERKINS, D. H., 1947, *Nature, Lond.*, **159**, 126.
 ROSSI, B., and GREISEN, K., 1941, *Rev. Mod. Phys.*, **13**, 240.
 VALLEY, G. E., 1947, *Phys. Rev.*, **72**, 772.
 WILLIAMS, E. J., 1938-1939, *Proc. Roy. Soc. A*, **169**, 531; 1940, *Phys. Rev.*, **58**, 292.

A Modification of Benoît's Method of Exact Fractions

By E. G. MICHAELIS

Physics Department, Birkbeck College, London

Communicated by R. Fürth; MS. received 3 February 1948, and in amended form 12 April 1948

ABSTRACT. A process is outlined by which the computations arising in the application of Benoît's Method of Exact Fractions to interferometric measurements can be shortened and simplified. A graphical illustration of the method is given, and it is compared with other modifications of Benoît's original technique.

§ 1. INTRODUCTION

WORK carried out in this department has shown that Benoît's "Method of Exact Fractions" as used in interferometry is capable of a slight modification, which has been found to shorten and facilitate the computations involved and which may also serve to extend the usefulness of the technique.

Benoît's method is commonly employed to find the order of interference of light between the plates of an interferometer from the approximately known distance between the plates and the measured fractional part of the order at the centre of the fringe-system. It was developed for use with the Michelson interferometer; nowadays it is widely used in work with the Fabry-Perot etalon and with such instruments as Kösters' interference comparator. Full accounts have been given by Benoît (1898), Childs (1926) and Williams (1930).

§2. THE METHOD OF EXACT FRACTIONS.

For the purpose of the method the integral part p and the fractional part f of the order of interference n at the centre of the fringe system are determined separately. An approximate value p_1 of p is found from the separation of the plates as determined by mechanical means and, in the case of the Fabry-Perot interferometer, f is obtained from the diameters of the interference rings. The true value of n_1 for a wavelength λ_1 is then given by

$$n_1 = p_1 + f_1 + m, \quad \dots\dots(1)$$

where m is a positive or negative integer to be determined.

Now m has to be so chosen that the fractional parts of the central orders n corresponding to wavelengths λ calculated by

$$n = n_1 \lambda_1 / \lambda \quad \dots\dots(2)$$

agree with those found experimentally. Usually m is given successive integral values until this desired agreement is achieved simultaneously for all wavelengths used. This method, though simple enough in theory, may nevertheless involve a large number of seven- or eight-figure computations when more than four wavelengths are used for the calibration, and the assumed p_1 differs appreciably from the actual value.

Methods for shortening this work have been developed by Kösters (1926) and by Pérard and Maudet (1927). Kösters describes a slide rule on which the fractional parts for various spectral lines are plotted against the separation, which may be read off when the required combination of fractional parts has been found. Pérard and Maudet show that, by a method to which the following is related, the plate-separation may be found in a series of approximations when the mere order of its magnitude is known.

§3. BASIS OF THE MODIFICATION

The method now suggested is based on the following consideration:—

For $m=0$ we have, by (1) and (2), $n = (p_1 + f_1) \lambda_1 / \lambda = p + \phi$, say, where p and ϕ are once more the integral and fractional parts respectively. For another value of m ,

$$n = (p_1 + f_1 + m) \lambda_1 / \lambda = p + \phi + m \lambda_1 / \lambda = (p + m) + \phi + m(\lambda_1 - \lambda) / \lambda.$$

If, therefore, the fractional orders f for the wavelengths λ are to agree with observation, then $\phi + m(\lambda_1 - \lambda) / \lambda$ may only differ from f by an integer, i. e.

$$m(\lambda_1 - \lambda) / \lambda = f - \phi + r, \quad r = 0, \pm 1, 2, \dots \quad \dots\dots(3)$$

or

$$m = (f - \phi + r) \lambda / (\lambda_1 - \lambda). \quad \dots\dots(4)$$

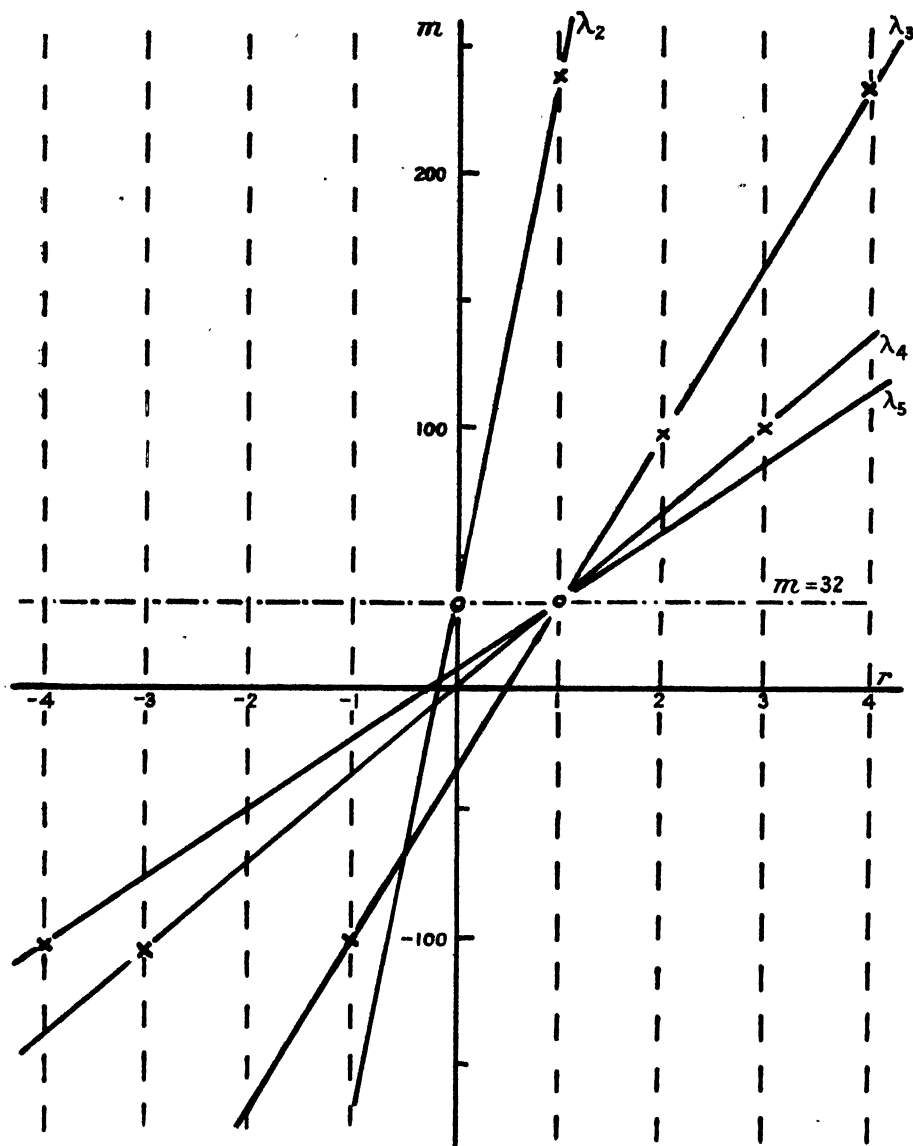
The integer m to be added to p_1 must of course be the same for all values of λ . Hence it is only necessary to solve (4) for a range of values of r for each wavelength until an m is found which is common to all. This will in general occur with a different r for each wavelength unless these lie fairly close together. Since m is a small integer, only a low degree of accuracy is required, and all auxiliary calculations may be carried out with a slide rule.

§4. GRAPHICAL INTERPRETATION

The method also permits of a graphical interpretation. By plotting m against r a straight line is obtained for each λ . The intersections of each line with $r = 0, \pm 1, 2, \dots$, represent values of m yielding the correct f for the corresponding

wavelength. The m to be selected for a general coincidence is then characterized by such points of intersection which for all λ 's lie on the same ordinate.

The diagram shows the straight lines thus obtained in the measurement of the plate-separation of a 1-cm. etalon, using neon light. The following wave-



× Partial near-coincidences.

○ General coincidence at $m=32$.

lengths were employed: $\lambda_1 = 5852.4878 \text{ \AA}$, $\lambda_2 = 5881.8950 \text{ \AA}$, $\lambda_3 = 5944.8342 \text{ \AA}$, $\lambda_4 = 6029.9971 \text{ \AA}$, $\lambda_5 = 6074.3377 \text{ \AA}$.

Near-coincidences for λ_1 , λ_4 and λ_5 occur at $m = -103$, and for λ_1 , λ_3 and λ_4 at $m = +99$, but the correct value for the simultaneous coincidence of all fractional parts is clearly given by $m = +32$.

§ 5. CONCLUSIONS

The method of calculating the correct value of m has been used by several students of this department and has been found to yield quick and reliable results. Using four wavelengths only eight seven-figure calculations are required, and it has been possible to carry out the entire work without the use of a calculating machine. Since the method is clearly capable of distinguishing the correct value of m from a range lying between ± 100 , the approximate separation of the plates need only be known to about $1/10$ mm. This fact might facilitate work where large plate-separations have to be used, e.g. in the determination of the metre.

The process outlined would appear to represent a final step in the method of Pérard and Maudet; it also seems the only one necessary unless the conditions are as stringent as those assumed by these authors. The method of finding m cannot, of course, compare in rapidity with that made possible by the use of Kösters' slide rule, but it will be recognized that the acquisition or the making of such an instrument will not be warranted unless a large number of routine measurements with a standard source are to be made.

ACKNOWLEDGMENTS

Thanks are due to Mr. R. E. Siday of this department and to Dr. H. J. J. Braddick of Manchester University for their kind help in the writing of this note. Especially valuable assistance was received from Mr. H. Barrell of the Metrology Division of the National Physical Laboratory, who kindly drew my attention to the work of Kösters and Pérard and Maudet.

REFERENCES

- BENOÎT, J. R., 1898, *J. Phys. Radium*, (3), 7, 57-68.
CHILDS, W. H. J., 1926, *J. Sci. Instrum.*, 3, 129.
KÖSTERS, W., 1926, *Z. Feinmechanik*, 34, 56.
PÉRARD, A., and MAUDET, L., 1927, *Comité International des Poids et Mesures, Travaux et Mémoires du Bureau*, 17, 35.
WILLIAMS, W. E., 1930, *Applications of Interferometry* (London: Methuen).

LETTERS TO THE EDITOR

An Isotopic Abundance Rule

F. C. Frank has recently (1948) drawn attention to the exceptions to the rule requiring even-odd nuclei to have at least one even-even neighbour of greater abundance. To the exceptions which he quotes (^{129}Xe , ^{147}Sm , ^{149}Sm , ^{195}Pt , ^{235}U) may be added ^{99}Ru , which is slightly more abundant than ^{100}Ru (Ewald 1943). We should like to comment on these exceptions.

^{99}Ru , ^{129}Xe . The abundances of these isotopes have almost certainly been enhanced by the decay, during geological time, of ^{99}Tc (half-life 4×10^6 years) and of ^{129}I , respectively. It seems likely (see Katcoff 1947) that ^{129}I has a sufficiently long half-life to have survived beyond the solidification of the earth's crust. Assuming that the excess ^{129}Xe content of the atmosphere is derived from ^{129}I originally in rocks which have by now suffered erosion, one can make a rough calculation of the half-life of ^{129}I . The value we obtain is about 4×10^8 years, in approximate agreement with the value suggested by Katcoff. Experiments.

are in progress here to check this hypothesis by examining geologically-old iodine-containing minerals for xenon by a micro-gas analytical technique. These two exceptions are in fact attributable to the unexpected β -instability of the parent isotopes, rather than to any irregularity on the part of the daughter-substances themselves.

^{147}Sm , ^{148}Sm , ^{235}U . These exceptions can readily be understood in terms of the α -instability of ^{148}Sm and ^{234}U . In these cases it is the decrease in the abundances of the neighbouring even-even isotopes, during geological time, which causes the anomaly.

^{196}Pt . No such explanation is available in this instance. It may be pointed out, however, that ^{196}Ir has not been reported; it may well be that it has a half-life too short to have survived to a significant extent, but still too long to give an observable activity in such bombardment experiments as have so far been carried out.

K. F. CHACKETT.

G. R. MARTIN.

Londonderry Laboratory for Radiochemistry,

University of Durham.

5th May 1948.

EWALD, 1943, Private communication to Flügge and Matthauch, *Ber. dtsch. Chem. Ges.*, **76**, 1.

FRANK, F. C., 1948, *Proc. Phys. Soc.*, **60**, 211.

KATCOFF, S., 1947, *Phys. Rev.*, **71**, 826.

Distribution Coefficients for the Calculation of Colours in the C.I.E. Trichromatic System for a Total Radiator at 2450° K.

Smith (1934) and Harding and Sisson (1947) have published tables of the distribution coefficients and the relative energy distributions for total radiators from 1500° to 3500° K. in steps of 250° and for the equal energy stimulus and standard illuminants A, B, C. These did not include data for 2450° K. required by us in connection with the estimation of chromaticity coordinates and integral transmission of colour filters.

The relative energy distribution and distribution coefficients for a total radiator at 2450° have therefore been computed in the same manner and for the same wavelength intervals as those prepared by Harding and Sisson. The use of "condensed" tables of this type having entries at every 0.01 micron reduces the time required in the calculation of chromaticity coordinates compared with that required for the use of the original C.I.E. tables which have twice the number of entries (Smith and Guild 1931-2). The previous tables have been calculated using a value of 14350 for the constant C_2 in the Planck formula, and this value was retained in the preparation of the tables here.

The general method followed is that given in a recent paper by Harding and Sisson (1947). Relative energy values were calculated from the Planck formula $E_{\lambda\theta} = C_1 \lambda^{-5} / \{\exp(C_2/\lambda\theta) - 1\}$ to seven decimal places, the constant C_1 being eliminated by calculating all energies relative to that at a wavelength of 0.56 micron. From these energy values the distribution coefficients were calculated using the distribution coefficients for an equal energy stimulus published by Smith (1934). The values were then adjusted so that the sum of the $(E_{\lambda}y_{\lambda})$ entries was 100 and the entries rounded off to four decimal places.

The relative energy distribution and the distribution coefficients are given in tables 1 and 2.

The accuracy of the tables was checked by calculating the chromaticity coordinates of a total radiator at 2450° K. and comparing the colour equation with published data (Harding 1944).

The figures were found to be in agreement, as is shown in the table below.

Origin	Colour equation ($C_2=14350$)
Calculated 7-figure tables	$0.4809261x + 0.4141740y + 0.1048998z$
Table 2	$0.480926x + 0.414174y + 0.104900z$
Published figures	$0.480926x + 0.414174y + 0.104900z$

Table 1. Relative energy values for a total radiator at 2450° K. calculated from Planck's equation: $C_1 = 14350$

Wavelength in microns	Relative energy	Wavelength in microns	Relative energy
0.38	4.901883	0.58	120.3480
0.39	6.391445	0.59	131.1161
0.40	8.197502	0.60	142.2398
0.41	10.35542	0.61	153.6855
0.42	12.89910	0.62	165.4198
0.43	15.86034	0.63	177.4074
0.44	19.26807	0.64	189.6137
0.45	23.14781	0.65	202.0026
0.46	27.52132	0.66	214.5389
0.47	32.40599	0.67	227.1876
0.48	37.81488	0.68	239.9144
0.49	43.75630	0.69	252.6852
0.50	50.23396	0.70	265.4676
0.51	57.24681	0.71	278.2301
0.52	64.78936	0.72	290.9424
0.53	72.85177	0.73	303.5756
0.54	81.42009	0.74	316.1022
0.55	90.47657	0.75	328.4958
0.56	100.0000	0.76	340.7331
0.57	109.9661	0.77	352.7905

Table 2. Distribution coefficients for total radiator at 2450° K.

Wavelength in microns	$E_{\lambda}\bar{x}_{\lambda}$	$E_{\lambda}\bar{y}_{\lambda}$	$E_{\lambda}\bar{z}_{\lambda}$	Wavelength in microns	$E_{\lambda}\bar{x}_{\lambda}$	$E_{\lambda}\bar{y}_{\lambda}$	$E_{\lambda}\bar{z}_{\lambda}$
0.38	0.0005	0.0000	0.0023	0.58	9.9237	9.4232	0.0173
0.39	0.0024	0.0001	0.0112	0.59	12.1188	8.9373	0.0136
0.40	0.0104	0.0003	0.0496	0.60	13.6079	8.0789	0.0102
0.41	0.0392	0.0011	0.1867	0.61	13.8791	6.9630	0.0048
0.42	0.1591	0.0048	0.7643	0.62	12.7058	5.6661	0.0022
0.43	0.4046	0.0166	1.9748	0.63	10.2816	4.2409	0.0000
0.44	0.6040	0.0397	3.0301	0.64	7.6341	2.9828	0.0000
0.45	0.7012	0.0793	3.6956	0.65	5.1653	1.9484	0.0000
0.46	0.7214	0.1483	4.1406	0.66	3.1789	1.1761	0.0000
0.47	0.5740	0.2660	3.7770	0.67	1.7946	0.6575	0.0000
0.48	0.3229	0.4718	2.7561	0.68	1.0010	0.3639	0.0000
0.49	0.1264	0.8197	1.8342	0.69	0.5198	0.1877	0.0000
0.50	0.0219	1.4636	1.2334	0.70	0.2689	0.0968	0.0000
0.51	0.0482	2.5965	0.8142	0.71	0.1465	0.0526	0.0000
0.52	0.3686	4.1328	0.4576	0.72	0.0746	0.0262	0.0000
0.53	1.0834	5.6541	0.2758	0.73	0.0383	0.0137	0.0000
0.54	2.1292	6.9912	0.1495	0.74	0.0199	0.0085	0.0000
0.55	3.5307	8.1064	0.0709	0.75	0.0089	0.0029	0.0000
0.56	5.3503	8.9552	0.0347	0.76	0.0046	0.0015	0.0000
0.57	7.5446	9.4245	0.0208	0.77	0.0016	0.0000	0.0000
				Totals:	116.1169	100.0000	25.3275
				Colour:	0.48093x + 0.41417y + 0.10490z		

Research Laboratories of
the General Electric Company Ltd.,
Wembley, Middlesex.
26th April 1948.

SYLVIA A. E. HIRD.
H. F. STEPHENSON.

HARDING, H. G. W., 1944, *Proc. Phys. Soc.*, **56**, 305.

HARDING, H. G. W., and Sisson, R. B., 1947, *Proc. Phys. Soc.*, **59**, 814.

SMITH, T., 1934, *Proc. Phys. Soc.*, **46**, 372.

SMITH, T., and GUILD, J., 1931-2, *Trans. Opt. Soc.*, **33**, 73.

REVIEWS OF BOOKS

Modern Physics, by G. E. M. JAUNCEY. Pp. xiv + 561. Third edition. (New York: D. van Nostrand Co., Inc., 1948.) 33s.

The fact that a book reaches a third edition in a comparatively short time shows that it has found favour amongst those to whom it was addressed. The task of a reviewer is therefore somewhat simplified, for he has only to point out possible improvements and note some of the small errors which escape the eye of even the most vigilant proof-reader.

It is noted that photographs of eminent physicists appear throughout the book but, unfortunately, the balance of selection appears to have been made in favour of American contemporaries: one wonders why Sir J. J. Thomson, Lord Rutherford, Mme Curie and Prof. Planck have been omitted? If it is a question of space, one must ask, what is the value of giving a list of Nobel prize-winners or of including Chapter II—"Some Useful Mathematics"? This chapter could easily be omitted from any future edition of the book, for there will undoubtedly be a cry for more space, and if a reader is not acquainted with the elementary calculus how can he be expected to appreciate, from the help given here, discussions on wave mechanics?

Now while it is the privilege of any author to use what symbols he likes to denote various physical quantities, provided, of course, that each is satisfactorily defined, yet in a book which is undoubtedly modern in outlook one is surprised to find that G is still used to denote the gravitational constant: why not γ , and let G be the strength of the field? It would also appear to be a retrograde step to use R as the symbol for the strength of an electric field. Usually E is preferred, but if this symbol is used for the charge on a positively ionized particle then X [cf. J. J. Thomson's *Elements*] could be adopted as the symbol: the use of R is particularly unfortunate as in the same book it is used to denote, among other things, the universal gas constant and an electrical resistance. The reviewer would like to see the symbols $(e/m)_-$ and $(e/m)_+$ used to denote the specific charge of an electron and a positively charged ion respectively when it is necessary to differentiate between them.

Again, should not the idea of "lines of force" be omitted from a book for senior students? If $\vec{B}$ is the magnetic induction, then $\int \vec{B} \cdot \hat{n} dS$, where $\hat{n}$ is a unit vector, etc., is the flux of $\vec{B}$ across S , and this is all that the phrase "lines of force" would seem to imply. We also find that the "gauss" is used as the unit of magnetic field strength and that the units for the universal gas constant are given wrongly: they should be $\text{erg.mole}^{-1}\text{deg}^{-1}$. In connection with units, a lack of consistency in style is noted: one would have thought that an American author would have followed the recommendations of Birge and not have written $c = 3 \times 10^{10}$ cm/sec. and $J = 4.18$ joules/cal. [p. 539], but $c = 3 \times 10^{10}$ cm. sec^{-1} and $J = 4.18$ joule.cal $^{-1}$. Also specific heat should have for its unit $\text{cal.gm}^{-1} \text{deg}^{-1} \text{C.}$; the definition of specific heat, on p. 165, is wrong, dimensionally.

The argument on p. 107 is very confusing. The author is endeavouring to calculate the capacitance per unit length of a cylindrical condenser so that since dV is always an increase in potential this must occur when an element of positive charge is brought nearer to a positive charge: in this instance $dV = -(2q/r) dr$ and $V = \int_b^a (-(2q/r) dr) = 2q \log(b/a)$.

There is still no discussion of the Zeeman effect: this omission was pointed out by a reviewer of the second edition, and several errors and misprints occur, the most amusing of which is: "In 1803 Rumford . . . married the widow of Lavoisier . . . Lavoisier had been guillotined in 1894 . . ." [p. 24]. One also finds that on p. 171 a cross-reference exists to matter which has been removed from the present edition. Finally, why retain the term "perfect gas" when an "ideal gas" is meant, and why print van der Waals with a capital "V"?

These are only small defects in what will still prove to be an exceedingly useful book to serious students of physics, but such errors are so common among writers and teachers that I have ventured to draw attention to them.

C. J. SMITH.

THE PROCEEDINGS OF THE PHYSICAL SOCIETY

VOL. 61, PART 3

1 September 1948

No. 345

A Theory of Transient Creep in Metals

By C. L. SMITH

Cavendish Laboratory, Cambridge

MS. received 25 March 1948

ABSTRACT. A new approach to the theory of transient creep in metals is described and the relation $\epsilon = \text{const. } T(1 - e^{-Ct})/t$ deduced, connecting the creep strain ϵ , the absolute temperature T and the time t . C is a constant.

This relation is compared with experimentally obtained curves.

§ 1. INTRODUCTION

FROM observations on the creep of non-alloyed metals at constant stress, Andrade (1911, 1914) found that the creep curves could be very satisfactorily represented by the formula

$$l = l_0(1 + \beta t^{1/3})e^{\alpha t}, \quad \dots\dots (1)$$

where l is the length of the specimen after time t , and l_0 the length at the moment $t=0$ when the sudden extension, occurring immediately after the application of the stress, is assumed to have ended. α and β are constants which depend on stress and temperature. If $\beta=0$, $l=l_0e^{\alpha t}$ and $(1/l)(dl/dt)=\alpha$.

With increasing time, the creep rate tends asymptotically towards the value given by this equation; at low temperatures α is practically zero. Andrade called this "viscous" flow; it is not Newtonian in that the strain rate $(1/l)(dl/dt)$ is not proportional to the stress. According to Becker (1925), such flow may occur by a process of thermally activated rearrangement of small numbers of atoms. To this category belongs recovery creep, which takes place at a constant rate such that there is equilibrium between strain hardening and thermal softening.

If the temperature is low relative to the melting point of the metal, $\alpha=0$ and $dl/dt = \frac{1}{3}l_0\beta t^{-2/3}$. The creep rate then tends asymptotically to approach zero with increasing time. This component of creep was called " β " flow by Andrade; Orowan (1947) has suggested the expression "transient creep" or "transient component of creep". It is this, the " β " flow, or transient creep, with which we shall be concerned in what follows.

Since Andrade's original work, transient creep has received little attention until recently, when several papers (Cottrell and Aytakin 1947, Los, unpublished, Mott and Nabarro 1948, Orowan 1947) have revived interest in the problem. In one of these, Orowan (1947) has suggested an expression for the creep rate by considering it as a thermally activated process, dependent on the form of the stress-strain curve; this is

$$d\epsilon/dt = (A/\epsilon^2) \exp(-B\epsilon^2 h^2/kT),$$

where $d\epsilon/dt$ is the creep rate, A and B are constants, h is the slope of the

yield stress-strain curve at the point under consideration, and ϵ is the extension measured from the point when the instantaneous extension, immediately following the application of the stress, is assumed to have ended.

In this expression the exponential term, $\exp(-B\epsilon^2 h^2/kT)$, expresses the probability of a successful thermal activation occurring; the term A/ϵ^2 is an empirical factor giving the length of the glide avalanche ensuing from one successful activation.

In what follows, a treatment of transient creep is suggested which, although very much simplified, does not contain such an empirical factor.

§ 2. THEORY

The immediate effect of the application of the load in a creep test is an extension which is practically instantaneous if the inertia of the testing machine is small. This is followed by rapid plastic flow, which decreases in velocity more or less rapidly, and finally goes over into slow plastic flow or creep. We assume that, at the instant when sudden extension is complete, there exists in the specimen a very large number of spots of stress concentration where the local stress is higher than the mean applied stress, but lower than the local yield stress. This will occur where dislocations become trapped at energy barriers, as, for example, at the sites of impurity atoms or crystallite boundaries. Each of these spots can be characterized by an activation energy E , such that a local stress fluctuation, caused by thermal agitation, will raise the stress in a volume V up to the yield stress and initiate glide at that point.

The significance of the volume V is that the stress fluctuation must occur in a volume large enough for the local glide process not to be annihilated by the elastic forces in the surroundings when the stress fluctuation has ceased. The energy required to produce this stress fluctuation is E , the activation energy. Since the stress concentration will vary from point to point, the specimen in a given state will be characterized by a relationship between the energy E and the number of spots $f(E)dE$ with activation energies between E and $(E+dE)$.

As plastic flow continues, the distribution function $f(E)$ changes. We must first investigate the form of the function describing this variation with time. Let the function at time t be $f(E, t) = f$; we denote $f(E, 0)$ briefly by f_0 . Consider only the spots with activation energies between E and $(E+dE)$ at time t : their number will be $f(E, t)dE = fdE$. During the ensuing short interval of time dt , some will be lost by thermal activation from which glide ensues; let this number be dN , where

$$dN = -C(f dE)e^{-E/kT} dt. \quad \dots\dots(2)$$

In addition, the glide which occurs during time dt will produce new spots between E and $(E+dE)$. The number of spots produced may depend on many factors, in particular on the past treatment of the specimen; since we do not know how it is influenced by these factors, we shall assume, for simplicity, that no new spots are created during creep.

With this assumption, equation (2) becomes

$$dN = d(f dE) = -C(f dE)e^{-E/kT} dt. \quad \dots\dots(2a)$$

Equation (2a) has a solution of the form

$$f = f_0 e^{-\lambda t}, \quad \text{where} \quad \lambda = C e^{-E/kT}, \quad \dots\dots(3)$$

or

$$f(E, t) = f(E, 0) e^{-\lambda t}. \quad \dots\dots(4)$$

If the average amount of glide resulting from the activation of a spot is a , then the glide in time dt from the spots with activation energies between E and $(E + dE)$ will be

$$aC(f_0 e^{-\lambda})e^{-E/kT} dE dt. \quad \dots\dots(5)$$

This follows by substituting the value of f from equation (4) into equation (2).

By integrating this expression over all possible energies E , we obtain for the creep rate $d\epsilon/dt$ the expression

$$\frac{d\epsilon}{dt} = aC \int_{E=0}^{E=\infty} (f_0 e^{-\lambda}) e^{-E/kT} dE. \quad \dots\dots(6)$$

The integral in (6) cannot be evaluated unless the function f_0 is known, and neither theoretical nor experimental work can, at the moment, give any indication of its form. If, however, a suitably simple form for the function $f(E, 0)$ is chosen, the integral (6) can be expressed by known functions. We shall assume, therefore, a distribution $f(E, 0) = \text{const.} = p$, such that there is an equal number of spots in equal energy intervals dE . The glide rate at time t , from equation (6), is then given by

$$\frac{d\epsilon}{dt} = aCp \int_0^{\infty} e^{-\lambda t} e^{-E/kT} dE. \quad \dots\dots(7)$$

Substituting $dE = kT(\lambda/C)e^{E/kT}$ (see equation (3)) in (7), we have

$$\frac{d\epsilon}{dt} = apkT \int_{E=0}^{E=\infty} e^{-\lambda t} d\lambda = apkT \left[\frac{e^{-\lambda t}}{-t} \right]_{\lambda=C}^{\lambda=0} = apkT \left\{ \frac{1 - e^{-Ct}}{t} \right\}. \quad \dots\dots(8)$$

It will be noticed that the limits of integration are taken from $E=0$ to $E=\infty$. Clearly, spots cannot have an activation energy greater than some maximum value $E_{\max}$; but since higher energy spots would make only a negligible contribution to the creep rate, the upper limit can be taken as $E_{\max} = \infty$ in order to simplify the final result.

§ 3. DISCUSSION

From equation (8) we see that the creep rate is proportional to (i) the absolute temperature T , and (ii) the function $F(t)$ defined by $F(t) = (1 - e^{-Ct})/t$.

In what follows we shall discuss only the function $F(t)$ describing the form of the transient creep curve at a constant temperature.

For values of t such that $Ct \ll 1$, $F(t)$ is a constant, since

$$F(t) = \frac{1 - e^{-Ct}}{t} = \frac{1 - (1 - Ct + C^2 t^2/2! \dots)}{t} = C + (\text{small terms}),$$

and the creep strain ϵ is, therefore, a linear function of the time $\epsilon = apkT(Ct)$.

When, however, $Ct \gg 1$, $F(t)$ approximates to $1/t$ and the creep strain

$$\epsilon = \text{const.} (apkT) \ln(\text{const. } t).$$

In figure 1 curves are plotted of $y = 1/Ct$ and also $y = (1 - e^{-Ct})/Ct$.

In the former case $y \rightarrow \infty$ as $t \rightarrow 0$, and in the latter $y \rightarrow 1$ as $t \rightarrow 0$. These curves show that the approximation $1/Ct$ fails when Ct is small, and it can be seen that for values of $Ct \lesssim 3$ the deviation is appreciable, but for values of $Ct \gtrsim 3$ the approximation is very good. To determine for what values of t the deviation is appreciable, it is necessary to know the value of C .

C has the character of a frequency, and by making certain assumptions Mott and Nabarro (1948) conclude that its order of magnitude should be 100–1000 times smaller than that of the frequency of the atomic oscillations in a lattice. They

deduce this by considering a precipitation-hardened metal containing line dislocations arrested at the sites of precipitate atoms which are assumed to be, on the average, about 200 Å. apart. The conclusion is that C is the frequency of the normal mode of oscillation of such dislocations 200 Å. long, from which is obtained a value for C of $10^8 - 10^9 \text{ sec}^{-1}$. This value will decrease with increasing distance between the points at which the dislocation is temporarily arrested, but even under the most favourable circumstances (i.e. with very high purity metals) it is difficult to see how an increase in the distance apart by a factor of more than 10^3 or 10^4 could be achieved. In this case C would become $10^4 - 10^5 \text{ sec}^{-1}$. From this we see that the approximation $F(t) = 1/t$ should be accurate for all values of t greater than about 10^{-4} sec .

By considering the movement of line dislocations over energy barriers, Mott and Nabarro also obtain a relation between creep strain and time

$$\epsilon = \text{const. } T^{2/3} (\ln Ct)^{2/3}.$$

This expression, of course, leads to $\epsilon \rightarrow -\infty$ for $t \rightarrow 0$; however, $t \simeq 1/C$ is the smallest value of t to which a physical sense can be attached.

A logarithmic relation between creep strain and time, $\epsilon = D \ln Et$, has already been suggested by several authors (Chevenard 1934, Phillips 1905, Tapsell and Prosser 1934), who have shown that it gives a good agreement with experimental results within the range of variables considered. In prolonged tests, however, it is found that the creep strain given by the above relation is less than that observed experimentally; Andrade's expression, on the other hand, under similar circumstances, would, in general, give too high values for the creep strain.

Examples of specimen creep curves are shown in figures 2-4.

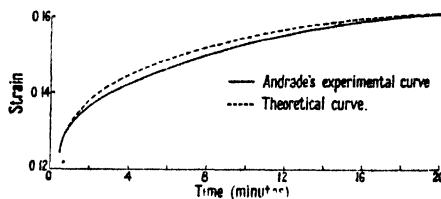


Figure 2.

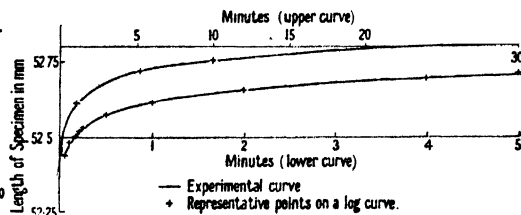


Figure 3.

The lower curve is the early part of the upper curve with an extended time scale.

The curve in figure 2 is taken from Andrade's paper (1914) and represents a creep test in which a lead specimen at -78°C . was subjected to a stress of 401 kg/cm^2 . At this temperature creep may be considered to be wholly transient.

Figure 3 shows a creep curve for copper* taken at room temperature at a constant stress of 1764 kg/cm^2 . The stress was kept constant by the use of a

* This was kindly supplied by J. Los (unpublished).

spring-loading testing machine in which the decrease in the cross-sectional area of the specimen was compensated by a reduction in applied load produced by the relaxation of the spring. In addition, in this experiment the inertia of the moving parts of the apparatus was kept as small as possible; it should therefore provide a reliable test for the theory, and it is seen that the logarithmic function does, in fact, provide a good fit. An attempt was also made to fit Andrade's expression $l = l_0(1 + \beta t^{1/3})e^{\alpha t}$ to this curve. It was found that this could only be achieved by giving α a small negative value, to which, of course, no physical interpretation can be attached. A very good fit was, however, obtained with $\alpha = 0$ and the exponent of the time 0.10 instead of 0.33 as above, i.e. $l = l_0(1 + \beta t^{0.1})$.

Finally, figure 4 shows the initial part of the flow curve for a zinc single crystal at room temperature. In this case there was without doubt a super-imposed component of viscous creep which was taken into account by an additional term proportional to t , i.e. $\epsilon = D \ln(Et) + Ft$.

With three unknown constants, the experimental and theoretical curves can be made to coincide at three points; these were chosen to be at $t = 3$ sec., $t = 16$ sec., and $t = 260$ sec. The agreement is again seen to be quite good. In fact, when it is remembered that the form of the distribution function $f(E, 0)$ was chosen arbitrarily, and purely from the point of view of simplicity in mathematical manipulation, the agreement between the theoretical and experimental results is unexpectedly good.

Before such a simple theory can be advanced further, it is necessary to know the form of the function $f(E, 0)$ more accurately, and also the manner in which this function changes as flow progresses. This would involve an investigation of the mechanism by which new spots are formed during creep. When this is accomplished, a theory on the lines outlined above will certainly give a more complete picture of transient creep.

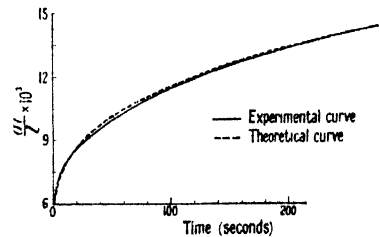


Figure 4. Zinc single crystal.

ACKNOWLEDGMENT

The author would like to thank Dr. E. Orowan for many helpful discussions, and the British Iron and Steel Research Association, under whose auspices the work was done, for financial assistance.

REFERENCES

- ANDRADE, E. N. DA C., 1911, *Proc. Roy. Soc. A*, **84**, 1; 1914, *Ibid.*, **90**, 329.
- BECKER, R., 1925, *Phys. Z.*, **26**, 919.
- CHEVENARD, P., 1934, *Revue de Metallurgie*, November, p. 473.
- COTTRELL, A. H., and AYTEKIN, V., 1947, *Nature, Lond.*, **160**, 328.
- MOTT, N. F., and NABARRO, F. R. N., 1948, *Report of Conference on Strength of Solids* (London: Physical Society), p. 1.
- OROWAN, E., 1947, *West of Scotland Iron and Steel Institute*.
- PHILLIPS, P., 1905, *Phil. Mag.*, **9**, 513.
- TAPSELL, H. J., and PROSSER, E. E., 1934, *Engineering*, **137**, 212.

The Thermal Conductivity of some Insulating Materials at Low Temperatures

By C. S. CHOW *

Imperial College, London

* Now returned to China

Communicated by Sir Alfred Egerton ; MS. received 21 May 1947, and in amended form 8 January 1948

ABSTRACT. The method of Lamb and Wilson modified for use at low temperatures is used for thermal conductivity measurements of insulating materials in the range of temperature $+18$ to -180°C . For finely granulated and loosely packed fibrous materials, the temperature gradient was experimentally determined and the intermediate values of the conductivity along the temperature gradient were obtained. For coarsely granulated or loosely packed materials, only the mean conductivity is observed.

The general shape of the conductivity-temperature curve in relation to the density of packing for various materials is discussed. For compressible fibrous insulators, the variation of conductivity with packing densities at low temperatures has been found similar to that at normal or moderate high temperatures. Granulated cork, vermiculite, cotton waste, sawdust and slag wool were amongst the materials investigated.

§ 1. INTRODUCTION

A CONSIDERABLE amount of work has been done on the thermal conductivity of common insulating materials both at normal and at high temperatures, but data concerning the conductivities at low temperatures are very limited.

The present investigation was carried out in the Department of Chemical Technology, Imperial College, during the early period of the war, in connection with work on the liquefaction of methane. Results relating to some useful insulating materials for a temperature range of about 10° to -180°C . were obtained and are here reported.

The plate method has been used by Ezer Griffiths (1928/9, 1932, 1936, 1940/41) and by Rowley, Jordan and Lander (1945) for determination of the heat conductivities of materials to -75°C . and -40°C . respectively and to -183°C . by Wilkes (1946). The spherical shell method has been used by Gröber (1909, 1910) and by Raisch and Weyh (1932) for measurements to -183°C .

The method used in the present investigation is based on that used at normal temperature by Lamb and Wilson (1899), but includes the introduction of a guard-ring arrangement. This method was chosen because it is easier to set up two concentric cylinders than to provide spherical shells. The insulating materials can be packed conveniently between the cylinders, and thermocouple elements can be fixed in precise positions to determine the temperature distribution along the heat flow path.

§ 2. APPARATUS AND METHOD OF EXPERIMENTS

In figure 1 is shown the design and arrangement of the apparatus. It consists of two concentric cylinders, the space between them being packed with the material under test. The inner cylinder was 2 in. in diameter, 19 in. long with 1/16 in.

wall thickness; it was maintained at a low temperature with a suitable cooling agent such as liquid oxygen and acted as the "axial cooler". The outer cylinder, which served both as "heater" and water flow calorimeter, was double-walled, 12.5 in. long by 10 in. internal diameter. The wall space was baffled so as to distribute the water flow evenly round the circumference.

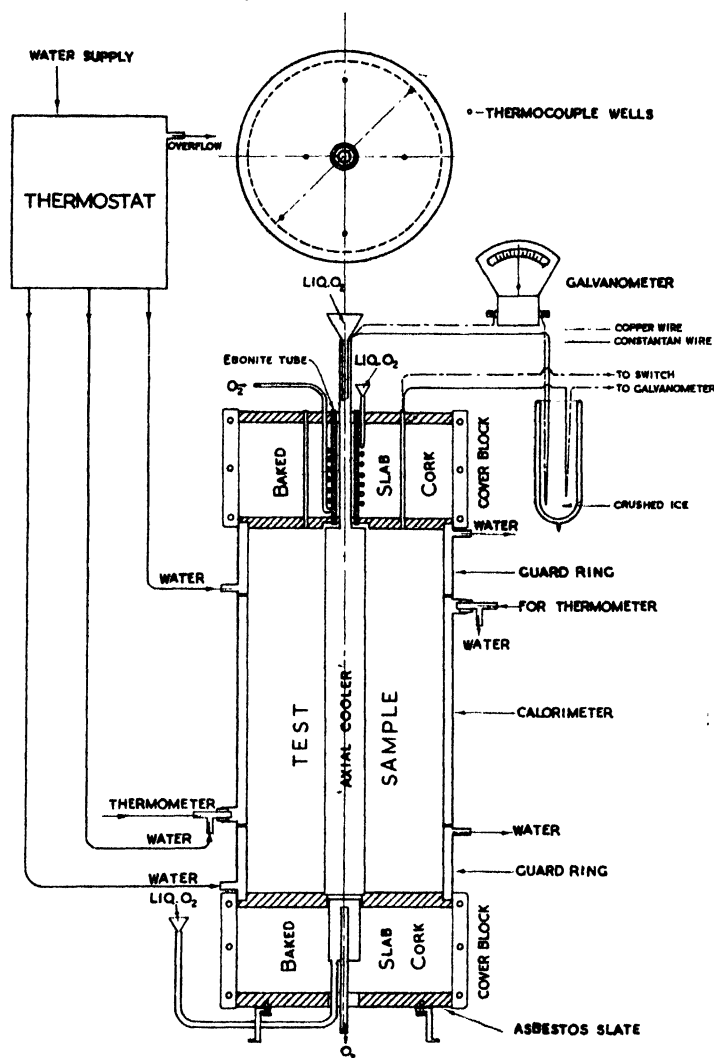


Figure 1. Apparatus.

Above and below the calorimeter were copper water jackets of the same diameter which served as "guard-rings", and the assembly was fitted at both ends with insulating cover-blocks which were equipped with auxiliary "axial coolers". Both guard-rings and cover-blocks were designed to eliminate "edge effects" from the heat flow, so that all the heat transferred from the calorimeter to the cooler could be assumed to flow at every point radially through the test material, i.e. in a direction perpendicular to the cylinder axis. The top cover-block was equipped at different radii with 3/16 in. holes parallel to the axis, which served as thermocouple wells.

Temperature measurements were made by means of copper-constantan thermocouples. The copper wire was 0.012 in. in diameter, and the constantan 0.007 in. The couples were calibrated at the normal boiling points of various pure cooling agents (liquid oxygen, methane, ethylene, dry ice etc.), and the readings thus obtained were then compared with the available data in the literature. Each thermocouple was sheathed in small straight glass-tubing and introduced through a "well" into the test sample, and movable in a direction parallel to the cylinder axis, so that the temperature along the vertical direction of the test sample could be determined for the corresponding radius. The temperatures at other radii were measured similarly with the aid of glass-sheathed thermocouples through the other wells. The temperature on each surface of the test sample was determined by soldering one thermojunction to the outside surface of the cylinder and another to the inner surface of the outer cylinder.

To conduct an experiment, the material under test was first carefully packed, attention being given to evenness of distribution in the case of fibrous materials. Then the whole apparatus was set up in the working condition, taking necessary precautions against the penetration of moisture through any interstices into the test sample. The thermocouples having been inserted, liquid oxygen was poured into the axial coolers, care being taken to avoid flooding which produced external cooling effects. The depth of cooling liquid in the main axial cooler was maintained constant (nearly full) with the aid of a thermocouple and galvanometer. Water from a thermostat at room temperature flowed continuously through both calorimeter and guard-rings, the rates of flow being made steady and equal. The temperature of the inlet and outlet to the calorimeter were measured with accurate mercury thermometers (read to within $\pm 0.01^\circ \text{C.}$), and the difference between them was adjusted to about 1°C. by regulating the rate of flow. Thus, the mean temperature of the calorimeter differed very little from room temperature and the tendency to transfer heat from the surroundings could be avoided. The effect of small variations in atmospheric temperature was minimized by wrapping both calorimeter and guard-rings with felt.

After a sufficient interval (generally from six to eight hours with a good insulator), the temperature drop of the calorimeter water or the thermocouple readings became steady, indicating that thermal equilibrium had been attained. The quantity of heat passing through the test sample was obtained from observations of both the flow rate and the temperature drop of the calorimeter water. A complete series of temperature observations was then taken, from which the temperature distribution along the path of heat flow was determined for the material under test. Finally, the assembly was dismantled and the test material weighed.

It may be noted that the method used for determining the temperature gradient as described above is only applicable to the conditions in which there is no difficulty in inserting the glass-sheathed thermocouples into the test sample. For the materials like cotton waste (in tangled thread form), this method cannot be applied. In such a case, only the temperature on each surface of the test sample can be taken; i.e. only a single value for the thermal conductivity (mean conductivity) is determined at a given temperature difference. For some materials which, in practice, are difficult to pack evenly, the method has also failed to give reliable results. However, in this apparatus, though only liquid oxygen was used, other cooling agents such as dry ice in toluene, if necessary,

can also be used for determining the conductivity values of the test material at other mean temperatures. It would thus be possible, by employing the same apparatus, to estimate the temperature coefficient of the thermal conductivity of those materials for which the temperature gradient is difficult to determine. With regard to those materials for which the temperature gradient can be determined with a reasonable degree of accuracy, the conductivity-temperature relationship for each material is readily derived from the experimental data at the given temperature difference as described in the next section.

§ 3. METHOD OF CALCULATION AND EXPERIMENTAL RESULTS

In the apparatus described the heat leakage from the ends is assumed to be negligible through the use of guard-rings, and the direction of heat flow is taken at all points to be radial through the section of the test sample under study. Then the cross-sectional area of the path is $2\pi rL$ at radius r (where L is the length of the calorimeter and r the distance from the cylinder axis) and the temperature at every point on such a cylindrical shell is constant. If the temperature gradient can be determined, the conductivity k is obtained from the formula,

$$k = \frac{q}{2\pi rL} \cdot \frac{1}{(dt/dr)}, \quad \dots\dots(1)$$

where q is the quantity of heat transferred in unit time, and dt/dr the temperature gradient (i.e. the slope of the temperature distribution curve, t vs r) at the point of radius r .

When the temperature gradient cannot be determined, then the mean conductivity k_m is obtained from the formula,

$$k_m = \frac{q}{2\pi L/\ln(r_2/r_1)} \cdot \frac{1}{t_1 - t_2}, \quad \dots\dots(2)$$

where t_1 and t_2 represent temperatures on the outside and inside surfaces of the test sample, and $2\pi L/\ln(r_2/r_1)$ is called the shape factor (Langmuir, Adams and Meikle 1913).

That these assumptions are justified may be illustrated by reference to the experimental results for finely granulated cork, packed to obtain a uniform bulk density. In figure 2(a)

are shown the observed longitudinal temperature readings on the cylindrical shell at several different radii. The first experiment was repeated after re-packing, when the bulk density differed by 0.1 lb/cu. ft. It can be seen that while there is no evidence of appreciable "edge effects" there are some slight irregularities in the temperature readings, the maximum deviation from the mean being

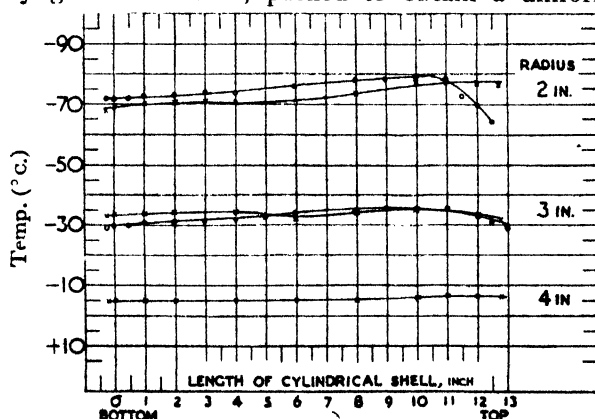


Figure 2(a). Longitudinal temperature readings.
Granulated slab cork.

Grain size : 49.5% 10-20 mesh and 59.5% through 20 mesh.
—○— for density of 6.3 lb/cu. ft.
—×— for density of 6.4 lb/cu. ft.

generally less than 10%. However, taking into account the small variations in grain size and the probable experimental error (for example, the inexact location of the thermojunctions), such a result is satisfactory. Taking the average height of ordinate for each curve to represent the temperature on the corresponding shell the results in table 1 are obtained. These are plotted in figure 3. The difference in the rate of heat transfer in the two experiments is only about 0.4 B.T.U./hour (less than 2%).

Table 1

Density lb/cu.ft.	Radius (inch)							Mean temp. of water (° c.)	Rate of flow Δt (° c.)	cm ³ /min.
	1	2	3	4	4½	4¾	5			
Temperature (° c.)										
6.3	-181	-75	-32.5	-5.0	—	—	14.5	17.0	0.73	131.0
6.4	-181	-74	-33.5	-5.5	7.0	11.0	14.5	16.5	0.80	121.4

By graphical differentiation of the t - r curve (figure 3) the temperature gradient against radius is obtained, and from these values the thermal conductivity at various temperatures is calculated by means of equation (1). The detailed results so obtained are set out in table 2, from which the conductivity-temperature curve is plotted (see figure 4(a)).

Table 2. Granulated slab cork (baked).

Density: 6.3-6.4 lb/cu. ft.
 $q=22.75$ B.T.U./hr.

Grain size { 49.5% 10-20 mesh.
50.5% through 20 mesh.

Radius r (inch)	t (° C.)	$\frac{dt}{dr}$ (° F./ft.)	$\frac{2\pi rL}{\text{(sq. ft.)}}$	Conductivity, k (from eq. (1))	
				B.T.U.*	C.G.S.†
1	-181	48.5×86.4	0.545	0.0099	0.000041
1½	-141	33.0 "	0.681	0.0117	0.000048
1¾	-113	23.8 "	0.818	0.0135	0.000056
1½	-91.5	18.5 "	0.954	0.0149	0.000062
2	-75	14.5 "	1.090	0.0166	0.000069
2½	-62	12.3 "	1.227	0.0174	0.000072
2½	-50.6	10.7 "	1.363	0.0180	0.000074
2¾	-40.8	9.3 "	1.499	0.0189	0.000078
3	-32.0	8.2 "	1.636	0.0196	0.000081
3½	-24.3	7.3 "	1.772	0.0203	0.000084
3½	-17.2	6.7 "	1.909	0.0206	0.000085
3¾	-10.8	6.1 "	2.045	0.0211	0.000087
4	-4.8	5.6 "	2.181	0.0215	0.000089
4½	0.7	5.2 "	2.317	0.0218	0.000090
4½	5.8	4.7 "	2.454	0.0227	0.000094
4¾	10.3	4.4 "	2.590	0.0230	0.000095
5	14.5	4.0 "	2.727	0.0241	0.000099

* Throughout this paper, k expressed in B.T.U. is in consistent units, i.e. B.T.U./(hr.) (sq. ft.) (° F./ft.).

† k expressed in C.G.S. units is gram-cal/(sec.) (cm²) (° C./cm.)

Conversion factor from B.T.U. to C.G.S.—multiply the former by 0.00413 to obtain the latter.

In the case of coarsely granulated materials the curves corresponding to figure 2(a) show wider deviations, the temperature being generally lower at the bottom and higher at the top of each cylindrical shell. Furthermore, for each corresponding shell the general shape of such curves depends largely on the

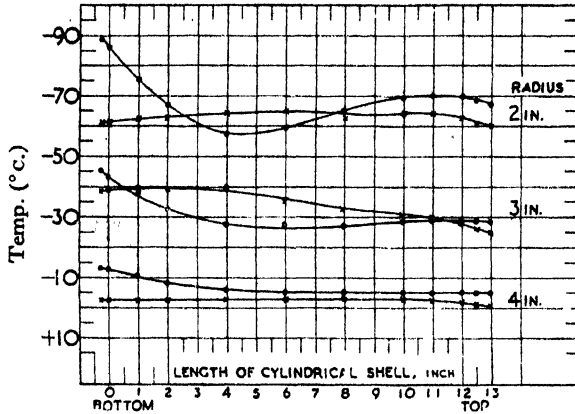


Figure 2 (b). Longitudinal temperature readings.
Slag wool.

—○— for density of 9.2 lb/cu. ft.
—x— for density of 12.3 lb/cu. ft.

relative proportions of the fractions of different grain size, and for the same packing the smaller the radius the greater are the deviations at the two ends of the shell.

For coarsely granulated materials as well as fibrous materials loosely packed, the curves such as shown in figure 2 (b) for slag wool at the density of 9.2 lb/cu. ft. are typical examples of the wide variations in temperature; the variations are generally much greater when the density of packing becomes still lower. In

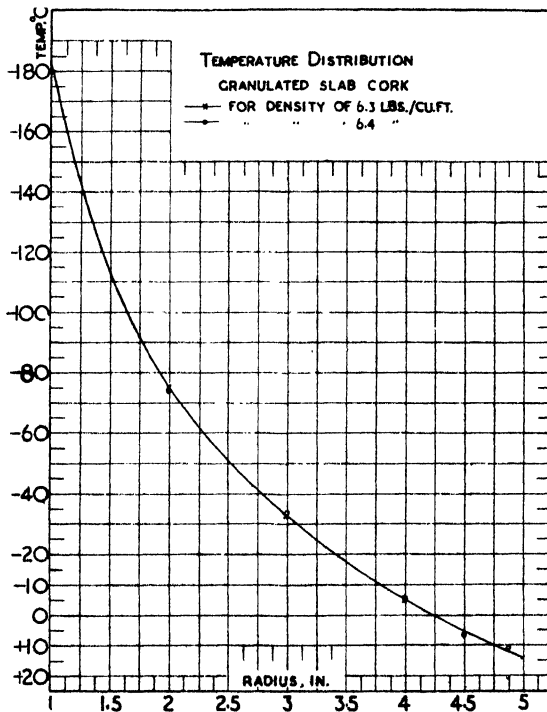


Figure 3.

such cases it is difficult to determine the temperature distributions with a satisfactory accuracy. Hence, for these materials it is only safe to calculate by means of equation (2) the mean values of the conductivity across the wide temperature range. The results so calculated are shown in table 4, including also the mean conductivity values for those materials given in table 3.

Table 3. Thermal conductivity at various instantaneous temperatures.

Temp. (° C.)	Units (see table 2)	No. of test materials (see descriptions in table 4)					
		1A	2A	3A	4	5	6
0	B.T.U.	0.0218	0.0380	0.0211	0.0358	0.0230	0.0224
	C.G.S.	0.000090	0.000157	0.000087	0.000148	0.000095	0.000093
- 40	B.T.U.	0.0190	0.0330	0.0183	0.0276	0.0187	0.0190
	C.G.S.	0.000079	0.000136	0.000076	0.000114	0.000077	0.000079
- 80	B.T.U.	0.0160	0.0275	0.0155	0.0225	0.0145	0.0155
	C.G.S.	0.000066	0.000114	0.000064	0.000093	0.000060	0.000064
-120	B.T.U.	0.0132	0.0227	0.0126	0.0192	0.0112	0.0120
	C.G.S.	0.000055	0.000094	0.000052	0.000079	0.000046	0.000050
-160	B.T.U.	0.0105	0.0191	0.0098	0.0170	0.0088	0.0089
	C.G.S.	0.000043	0.000079	0.000040	0.000070	0.000036	0.000037
-200*	B.T.U.	0.0085	0.0163	0.0070	0.0157	0.0062	0.0064
	C.G.S.	0.000035	0.000067	0.000029	0.000065	0.000026	0.000026

* Values extrapolated.

Table 4. Mean thermal conductivity
(Cold face temperature -181° C.)

No.	Material	Description	Bulk density lb/cu. ft.	Mean temp. (° C.)	Mean thermal conductivity	
					B.T.U.(1)	C.G.S.(2)
1A	Granulated cork (baked)	{ 49.5% 10-20 mesh 50.5% through 20 mesh }	6.3	-82.5	0.0159	0.000066
1A	do.	Repeated after re-packing	6.4	-82.7	0.0162	0.000067
1B	do.	4-10 mesh	5.4	-83.4	0.0168	0.000070
2A	Vermiculite*	{ 10-14 mesh 4-10 mesh }	13.5	-86.2	0.0271	0.000112
2B	do.	{ 7.9% 3-4 mesh 26.5% 4-10 mesh 65.6% 10-14 mesh }	9.0	-86.0	0.0199	0.000082
2C	Vermiculite (mixed)	{ 7.9% 3-4 mesh 26.5% 4-10 mesh 65.6% 10-14 mesh }	9.8	-85.7	0.0244	0.000101
3A	Sea-weed product	Powder form	8.0	-82.3	0.0155	0.000064
3B	do.	{ 50.0% block 49.5% powder }	7.3	-83.2	0.0201	0.000083
4	Sawdust (air-dried)	{ 4.0% on 10 mesh 24.4% 10-20 mesh 61.6% 20-40 mesh 10.0% through 40 mesh }	15.5	-82.3	0.0243	0.000100
5	Wadding	Crude cotton-wool	2.6	-83.4	0.0149	0.000062
6A	Slag wool		12.3	-82.3	0.0150	0.000062
6B	do.		9.2	-85.9	0.0115	0.000048
6C	do.		8.1	-84.9	0.0107	0.000044
6D	do.		7.0	-84.6	0.0113	0.000047
6E	do.		5.9	-83.9	0.0125	0.000052
6F	do.		5.0	-83.0	0.0134	0.000055
6G	do.		3.7	-83.0	0.0183	0.000076
7	Cotton waste	Tangled thread form	8.2	-86.7	0.0232	0.000092

* Vermiculite (in the expanded form) was received from the importers ground and graded 3-14 mesh. It was tested as received and then sieved so as to give two fractions, 4-10 and 10-14 mesh. These fractions were examined separately.

† Sawdust was a mixture of soft pine with a small amount of oak.

(1) B.T.U. per hour per sq. ft. for 1 ft. thickness and 1° F. temperature difference.

(2) Gm-cal. per sec. per cm^2 for 1 cm. thickness and 1° C. temperature difference.

For fibrous materials loosely packed the shape of the curves is particularly dependent on the uniformity of packing. This is illustrated by the curves in figure 2(b) for slag wool at two packing densities. The curves for the lower density (9.2 lb/cu. ft.) indicate the effect of compression of the bottom layer material under its own weight as well as that of uneven distribution of the packing material, while for much higher density of packing (12.3 lb/cu. ft.) very slight irregularities in the temperature readings are shown, with the one exception of the curve for 3 in. radius which is probably due to the effect of uneven distribution.

In general, for fibrous materials tightly packed such as slag wool at the density of 12.3 lb/cu. ft. and for materials finely granulated or in the form of powder, the variations in temperature for each cylindrical shell are very small, and therefore the temperature distribution along the path of flow can be determined with a

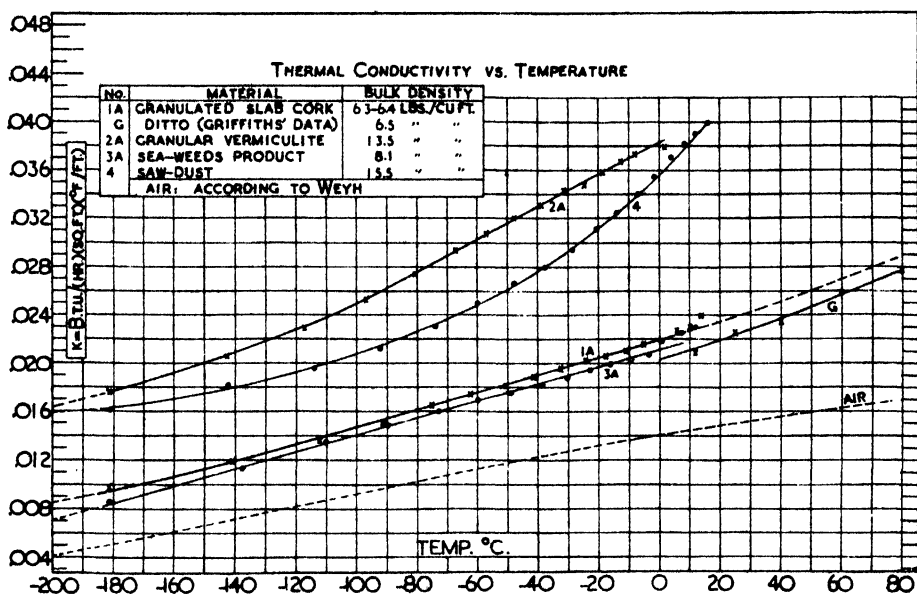


Figure 4 (a).

reasonable degree of accuracy. The conductivity-temperature relationships are derived by the method described above, the results so obtained being shown in figure 4(b). The results at several temperatures as taken from these smoothed curves are presented in table 3. Above 0° c. values of the conductivity are less certain than at lower temperatures, since in this region the variation of temperature gradient with path is small and consequently the influence of errors in temperature measurement on the calculated results is considerably increased.

The conductivity-temperature curve is also given for sawdust. This material consists of a large fraction of small particles but only a very small fraction of large particles or "chips", and, as the whole mass of the test sample can be evenly distributed, the packing density is fairly uniform. The curves corresponding to figure 2 show no particular irregularities; the temperature gradient is therefore determined without any great uncertainty.

§4. DISCUSSION OF RESULTS

As shown in figure 4 (a), the conductivity-temperature relationship is approximately linear over a wide range of temperature when insulating materials tested are closely packed to a uniform density (Nos. 1A, 3A, 6A etc.). However, when materials tested are not closely packed to give a uniform density, the variation of conductivity with temperature seems to deviate from the linear relationship, the degree of deviation being dependent on the tightness and uniformity of packing (cf. No. 6 and No. 5) (Rowley *et al.* 1945). The conductivity-temperature curve is most concave upwards for material most loosely packed.

The results indicated by figures 4 (a) and (b) can be compared with those shown in table 4 at the same temperatures; when k varies linearly with t (see curves 1A, 3A, 6A), the experimental value of k_m is correctly represented at t_m , no matter what the temperature drop across the test sample may be. However, if k is not a linear function of t , the two results show appreciable difference from each other, this difference being dependent on the degree of deviation from a linear relationship of the k - t curve. For instance, in the case of sawdust the difference between the experimental and the instantaneous values of the conductivity at -82.3°C . is about 2.5 times as great as that for wadding at -83.4°C .

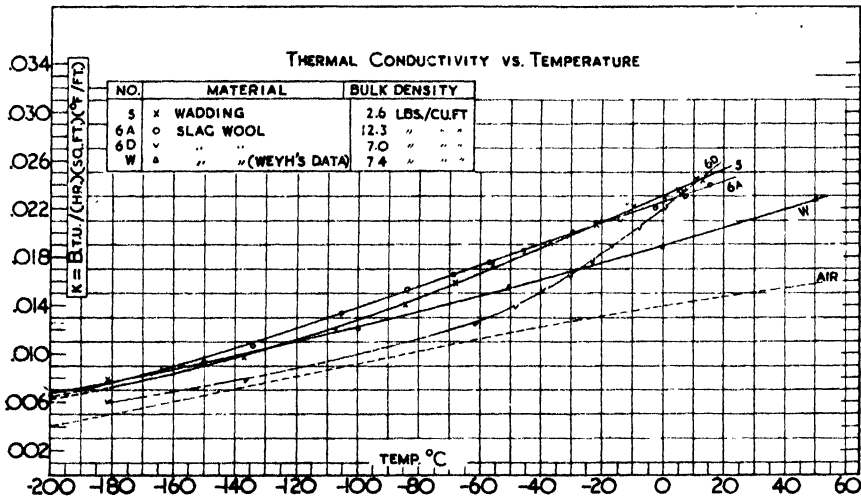


Figure 4 (b).

In a comparison of the conductivity data for a given material obtained by different methods, it is of importance to ascertain that the material tested by different investigators not only has the same properties (such as source of raw material, method of manufacture, density etc.) but also is tested in the same temperature range.

For slag wool the results reported by Raisch and Weyh (1932) give quite fair agreement with the present data in the low temperature range, but the mean values across wide temperature ranges cannot in fact represent the true behaviour of the conductivity for loosely packed materials.

For finely granulated slab cork, the conductivity was found to be approximately linear with temperature, and the extrapolation of its k - t curve (No. 1A) in the direction of higher temperature yields figures which compare well with

Griffiths' results (1940/41) for the same material at 6.5 lb/cu. ft. as shown in figure 4(a) (No. G).

For compressible materials the thermal conductivity is a function of the density of packing at a given temperature, and for each material there is an "optimum" density which corresponds to a minimum conductivity (20) (Griffiths 1940/41). In table 4 it is seen that the mean conductivity for slag wool at about -84°C . has a minimum value at about 8 lb/cu. ft. packing density, which is in agreement with Griffiths' results obtained in a 20 in. apparatus (1928/29, 1932, 1936, 1940/41).

It is also clear that, for a given granular material the conductivity depends on grain size, but the variation of conductivity with grain size may be different in nature for a different kind of granular material within the same range of grain size. Thus, in table 4, the apparent mean conductivity for granulated cork becomes higher when the test sample consists of comparatively large grains (4–10 mesh), while for granular vermiculite the conductivity becomes still lower as the proportion of the large grains (4–10 mesh) present is higher. This is probably connected with the porosity and other physical properties of the materials tested.

As already pointed out by various investigators, the amount of heat transferred through an insulating material is dependent on the conductivity of the material itself, and the transfer through air cells or interspaces within the pack is not only by conduction but also by radiation and convection. The relative amount of heat transferred by each of these processes is associated with the density of packing and the grain size of the material. It is generally agreed that, at higher densities of packing the conduction of heat by the material itself plays an increasingly important rôle, but at low densities or in a coarsely granulated material there are different views on the mechanism of the heat transfer through air spaces. According to Griffiths, convection takes place in the air spaces as observed in experiments on granular materials with a hot plate in a vertical position. Other authors suggest radiation as an important factor (Rowley *et al.* 1945), which causes the increase in slope of the curve.

In the present experiments, decrease of bulk density with the height occurred in loosely packed and coarsely granulated materials; consequently the thermal resistivity near the bottom with comparatively large densities was found to be appreciably greater than that towards the top, as can be illustrated by the curves for the longitudinal temperature readings (figure 2(b)). This behaviour is rather different from that observed by Griffiths. Furthermore, the k - t curve obtained was also seen to be most concave upward for the most loosely packed material.

When the effect of radiation and convection is reduced to a minimum, the apparent conductivity of an insulating material is obviously dependent on the conductivity of the material itself and of air in the interspaces. For materials of even porous structure the effect due to any radiation or convection may be negligibly small, so that the difference between the apparent conductivities of such materials could be taken to indicate the difference between the true conductivities of the solid part of the test samples. If such materials are also similar in physical structure, then the general shape of their k - t curves would be similar to each other, as can be seen from the apparent similarity of the two curves (figure 4(a)) for finely granulated cork (No. 1A) and powdered seaweeds product

(No. 3A). In line with the same view, the k - t curve for wadding (figure 4(b)), though slightly different from that for slag wool at 12.3 lb/cu. ft., might become similar to the latter if the wadding were packed to the same degree of tightness and evenness.

Since most thermal insulators consist of porous or granular materials, the conductivity of air is regarded as the possible minimum value. The k - t curve for air according to Raisch and Weyh (1932) is shown in figure 4 for comparison. The materials tested with the apparent conductivity values nearest to the conductivity of air are granular baked slab cork, powdered seaweeds product, slag wool and wadding.

It must be pointed out that the materials tested with comparatively low conductivity values are not all useful for low temperature work. In selecting an insulation for practical applications, there are other important factors which must be taken into consideration, such as moisture resistance, mechanical and thermal strength, density, and cost (cf. Quarmby 1942).

ACKNOWLEDGMENTS

The author wishes to record his indebtedness to Prof. Sir A. C. Egerton, Sec.R.S., Head of the Department of Chemical Engineering and Applied Chemistry, Imperial College, for continued interest and encouragement; to Prof. D. M. Newitt, F.R.S., for helpful advice on the design of the apparatus, and to Dr. J. H. Burgoyne for assistance in the work. Finally, the author is particularly indebted to the Universities' China Committee in London for a grant which made possible the carrying out of the work.

REFERENCES

- EGERTON, A., and PEARCE, M., 1945, *J. Inst. Fuel*, **18**, 168 (see refs. cited).
 GRIFFITHS, E., 1928/9, *Proc. Phys. Soc.*, **41**, 151.; 1932, *Structural Energy*, **10**, 202 ; 1936, *Chem. Engng. Congr.*, **1**, 366 ; 1940/41, *Proc. Brit. Assocn. Refrig.*, **37**, 15.
 GRIFFITHS, E., and AWBERY, J. H., 1935, *Proc. Brit. Assocn. Refrig.*, **31**, 93.
 GRIFFITHS, E., POWELL, R. W., and HICKMAN, M. J., 1942, *J. Inst. Fuel*, March.
 GRÖBER, H., 1909, *Z. ges. Kalte., Ind.*, **16**, 84 ; 1910, *Z. Ver. dtsh. Ing.*, **54**, 1319.
 LAMB, C. G., and WILSON, W. G., 1899, *Proc. Roy. Soc.*, **65**, 285.
 LANGMUIR, I., ADAMS, E. Q., and MEIKLE, G. S., 1913, *Trans. Amer. Electrochem. Soc.*, **24**, 55.
 QUARMBY, B., 1942, *British Plastics*, **13**, 376 ; **14**, 32.
 RAISCH, E., and WEYH, W., 1932, *Z. ges. Kalte., Ind.*, **39**, 123.
 ROWLEY, F. B., JORDAN, R. C., and LANDER, C., 1945, *Refrig. Engng.*, **50**, 541.
 WILKES, J., 1946, *Refrig. Engng.*, **52**, 37.

The Emission of Negative Ions from Oxide Coated Cathodes

By R. H. SLOANE AND C. S. WATT *

Queen's University, Belfast

* Now at Cavendish Laboratory, Cambridge

MS. received 1 December 1947

ABSTRACT. The negative ions emitted from an oxide coated cathode and nearby electrodes have been analysed, using both a demountable metal mass spectrometer and glass sealed-off tubes to which normal cathode-ray tube high-vacuum technique could be applied. The mass-to-charge ratios of a large number of negative ions from the different electrodes have been determined. Retarding potential measurements have enabled energy distribution curves to be plotted for the more intense ion beams, and permit of discrimination between negative ions produced thermionically and those produced by bombardment. Some observations have been made on the dependence of the ionic thermal emission on the cathode temperature, and the variation of this emission with time after changes in cathode temperature, accelerating field and cathode surface conditions. The results are discussed and compared with those of other workers in this field.

§ 1. INTRODUCTION

IT has long been known that the so-called "black spot" which develops on the screen of some high-vacuum cathode-ray tubes is produced by negatively charged heavy particles (von Ardenne 1935, Levy and West 1936) which are projected from the cathode region and focused on to the screen by the electrostatic lens. Bachman and Carnahan (1938) and Broadway and Pearce (1939) have been able to produce radial lines of black spots or negative ion mass spectra on the screen by applying a strong transverse magnetic field in the neighbourhood of the electron gun. More recently a paper by Schaefer and Walcher (1943) has become available describing results which were obtained by using a cathode-ray tube gun followed by a mass spectrometer with 48° deflection and an electrometer detector. Schaefer and Walcher, however, confined their attention to a demountable apparatus using grease sealed ground joints. The work described in the present paper is a comprehensive analysis of the negative ions from oxide coated cathodes and neighbouring electrodes, made with both a demountable metal apparatus and glass sealed-off tubes to which normal cathode-ray tube high-vacuum technique could be applied. Each complete apparatus constituted a mass spectrometer of the Dempster type, and permitted definite identification of the mass-to-charge ratios of the several negative ions observed, as well as determination of the energy distribution of the more abundant ions. It was also possible to investigate the dependence of the ionic thermal emission on the cathode temperature, and the variation of this ionic emission with time after changes in cathode temperature, accelerating field and cathode surface conditions. The experimental work was done between 1938 and 1940 but, apart from a preliminary note (Sloane and Cathcart 1939), publication in detail has had to be postponed on account of the war.

The general scope of the investigation is : (a) It makes an addition to the comparatively scanty information available concerning the mass numbers of atoms and radicles which form stable negative ions. (b) It provides new data for the emission from oxide coated cathodes. (c) It gives new information in

connection with the still incompletely understood phenomenon of cathodic sputtering. (d) It has a bearing on the design of cathode-ray tubes (Sharpe 1946, Bachman 1945).

The apparatus employed is described in §2 and the main results obtained are recorded in §3. Our results are discussed and compared with those of the other workers in this field in §4.

§2. APPARATUS

(i) Sealed-off glass tubes

Two sealed-off glass tubes were used. The essential features of the first are shown in figure 1. It included a negative ion source I, a magnetic deflection chamber D and a receiving chamber R. The source consisted of an indirectly heated oxide coated cathode 8 mm. long, with its heater connected to F_1 and F_2 , and its emitting surface connected to C. The cathode was surrounded by a cylindrical anode 1.0 cm. in diameter, which was connected to G_1 . A slit (1 cm. $\times$ 3 mm.) in the lower side of the anode was covered with tantalum gauze made from 36 gauge wires spaced 0.5 mm. apart, which will be referred to as the first gauze. A second similar tantalum gauze connected to G_2 was placed between the first gauze and the source slit (8 mm. $\times$ 0.75 mm.), the latter being connected to S_1 . All the electrodes, except the cathode and the two gauzes, were made of nickel-chrome which had previously been degassed in a vacuum furnace. G_2 was usually connected to S_1 . The inner surface of D was coated with colloidal graphite ("Aquadag") which gave both good electrostatic screening and a rough surface which reduced reflection of the ions at the wall to a minimum.

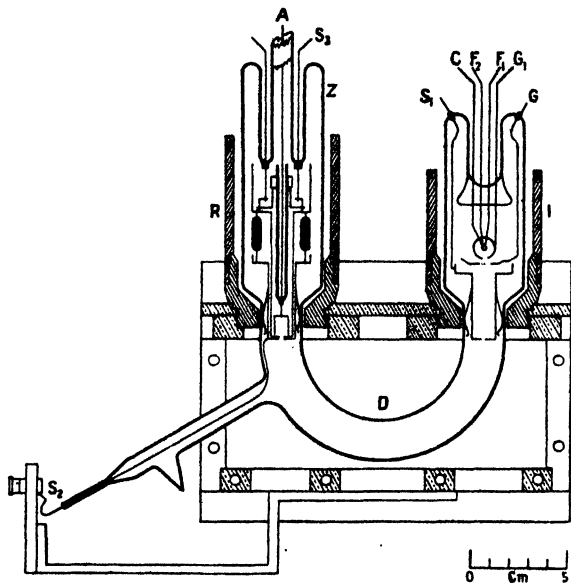


Figure 1. First glass sealed-off tube and mounting. In the tube, glass is shown by heavy lines, metal by lighter lines. The iron magnetic screen is shaded with full lines, the brass supports with alternate broken lines.

The Faraday cylinder, connected to A, was provided with guard rings and an electrostatic screen carrying the collecting slit, which was connected to S_3 . This screen was connected electrically to another screen (not shown in the diagram) round the outside of the glass envelope Z. This in turn formed part of the electrostatic screen for the lead from the Faraday cylinder to the grid of an electrometer valve (Osram Type T) used to record the negative ion current. In this way the Faraday cylinder and its leads were screened completely except for the small opening of the collecting slit S_3 (6 mm. $\times$ 0.75 mm.). A second wider slit (9 mm. $\times$ 2.6 mm.) connected to S_2 was placed in front of the collecting slit. Contact was made with the Aquadag by means of spring strips as shown,

and, in order to eliminate any possible potential drop along the Aquadag between the source and the collecting slits, S_1 was connected metallically to S_2 .

After assembly the tube was degassed for several hours in a furnace under vacuum. The electrodes in the source side were further degassed by eddy current heating, and the cathode activated in the usual way. The tube was then sealed-off and mounted in the same electromagnet as was used with the metal mass spectrometer (§2(ii)). The pole pieces were spaced by brass pillars which supported iron magnetic screens surrounding the source and detector chambers as shown in figure 1.

The negative ion accelerating voltage (V_n) was maintained steady at up to 2800 volts by a chain of neon tube stabilizers (Cossor Type S130). A potentiometer connected across a pre-selected section of this chain allowed fine adjustment of V_n .

The second tube, of which less use was made since it had a less good vacuum, had a somewhat similar electrode system, but with G_2 omitted. Advantage was taken of the small quantity of residual gas for some experiments described in §3(vii). Even in this tube the gas pressure was certainly not more than 10^{-4} mm. Hg and probably not more than 10^{-5} mm.

(ii) *The metal demountable system*

This mass spectrometer was the one previously described by Sloane and Press (1938b). It had heavy pole pieces spaced 9 mm. apart by a vacuum tight nickel-plated brass rectangular frame, and was constructed to permit of simultaneous deflection of two beams through 180° . In the present work only the larger of the two arcs was used and some slight structural modifications were made. The electrode systems were somewhat similar to those in the glass tubes. On the source side there was an oxide coated cathode C with a nickel backing plate at the same potential, a single gauze G and a slit S_1 (7 mm. $\times$ 0.75 mm.), while on the detecting side the Faraday cylinder was mounted behind two slits S_2 (8 mm. $\times$ 3 mm.), and S_3 (9 mm. $\times$ 0.75 mm.), S_3 being in the electrostatic screen surrounding the Faraday cylinder, as in the sealed-off tube. Both the source and detecting chambers were magnetically screened. All joints were made vacuum tight with a grease of low vapour pressure (Apiezon L).

§ 3. EXPERIMENTAL RESULTS

Mass spectra were recorded in all cases by keeping the magnetic deflecting field constant and applying a variable voltage V_n between C and S_1 , so as to accelerate negative ions towards S_1 . By varying V_n , one negative ion beam after another was brought through the slits into the Faraday cylinder. A voltage V_1 , accelerating negative ions in the same direction, was applied between the cathode C and the gauze G in the metal apparatus and second glass sealed-off tube, and between C and G_1 in the first glass sealed-off tube.

The resolving power of each apparatus was arranged so as to separate all peaks of interest without unnecessarily reducing the currents to be measured, which it was convenient to keep greater than 10^{-15} amp.

(i) *Ion spectra : metal demountable system*

Two mass spectra, typical of a large number obtained using the demountable metal system, are shown in figure 2. These were taken after a new cathode

had been activated and operated under normal conditions for five hours. The spectrum (a) was obtained with a total cathode emission of 2.4 ma., while spectrum (b) was obtained with the cathode temperature a little higher and the emission varying slowly from 3.1 ma. to 4 ma. during observations in which the light ions were recorded first. V_1 was 100 v. for spectrum (a) and 200 v. for spectrum (b). The voltage scale along the top of the diagram gives V_n .

Since V_1 was applied between C and G, while the main variable accelerating voltage V_n was applied between C and S_1 , it follows that V_n and $V_n - V_1$, together with any initial energy possessed, determined the energy with which negative ions liberated from the cathode and gauze respectively passed downward into the deflecting chamber D. An ion from the gauze would, therefore, enter the

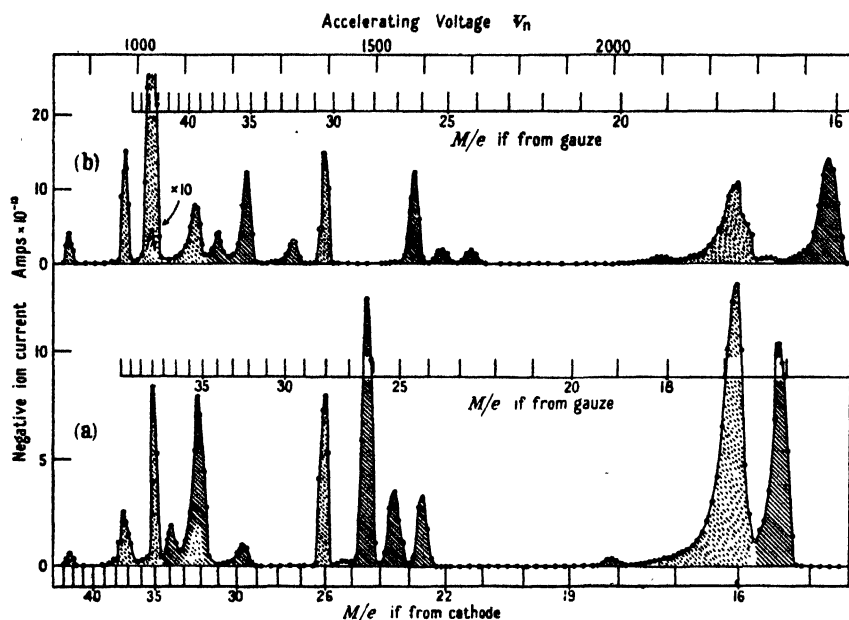


Figure 2. Negative ion mass spectra taken with the metal demountable apparatus. Peaks shaded with broken lines are due to ions from the cathode, those shaded with full lines to ions from the gauze.

(a) $V_1 = 100$ volts.

(b) $V_1 = 200$ volts.

Faraday cylinder at a value of V_n greater by an amount V_1 than the value in the case of a similar ion from the cathode. If V_1 were increased and a new spectrum taken, peaks due to ions from the gauze would move to higher values of V_n , while those from the cathode would remain fixed. An example of the resulting displacement is seen in figure 2. This method was used to determine the place of origin of the different ions throughout all of our work. The peaks shaded with full lines in figure 2 and elsewhere are due to ions from the gauze, and those shaded with broken lines to ions from the cathode. Where variation of V_1 revealed overlapping of two peaks from gauze and cathode respectively, the approximate relative intensities are indicated by double shading. Where ion beams were particularly intense both the base and, on a reduced scale, the peak are shown. The mass-to-charge ratio (M/e) scales in figure 2 have been calculated by taking the position of the ion with $M/e = 26$ in atomic mass units, from the cathode, as standard (see § 3(ii)), and staggering the scales for ions from the

gauze by an amount V_1 . The current scale for spectrum (b) is smaller than for spectrum (a). The peaks at mass numbers 42, 35 and 37 from the cathode have increased appreciably in intensity with increase in V_1 , while that at mass number 26 has increased slightly. Peaks at mass numbers 32 from the cathode and 35 from the gauze are superimposed in (a) but separated in (b).

No ions of mass number greater than 42 were found in these cases although search was made for them, and the available accelerating voltage was not high enough to allow a search for ions lighter than 16. Using a lower value of magnetic field a beam of small intensity at mass number 12 was observed.

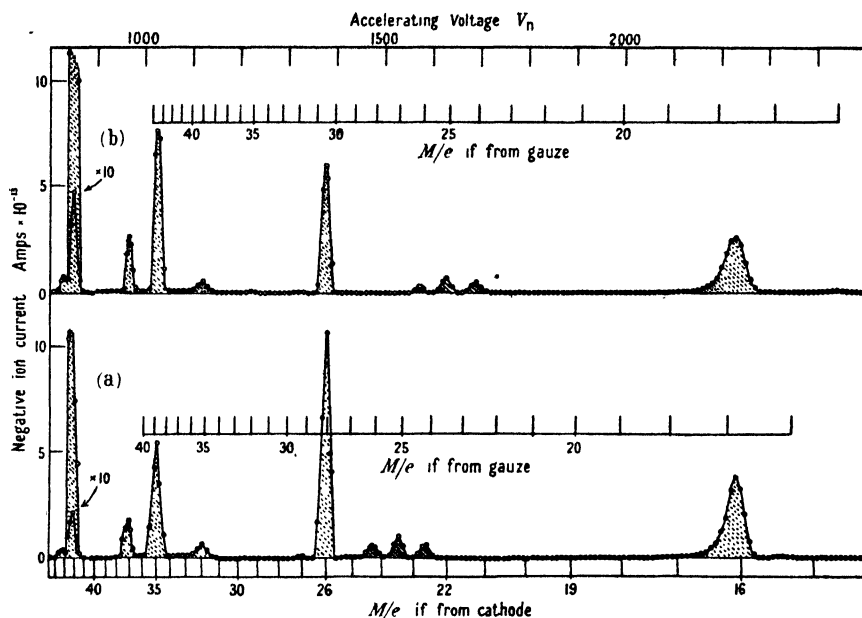


Figure 3. Negative ion mass spectra taken three days after those shown in figure 2, a partial vacuum having been maintained in the meantime. The shading of the peaks is the same as in figure 2.

(a) $V_1=100$ volts.

(b) $V_1=200$ volts.

Figure 3 shows two spectra taken three days after those shown in figure 2. In the interval a partial vacuum had been maintained (pressure about 10^{-3} mm. Hg) with the cathode cold and all accelerating voltages reduced to zero. An attempt was made to reform the cathode, which was then aged for three and a half hours before readings were commenced for spectrum (a) with V_1 100 v. The second spectrum (b) was taken with V_1 200 v. The cathode emission varied from about 3 ma. at high values of V_n to about 1.5 ma. at low values of V_n in each case. The relative heights of the peaks are quite different from those in figure 2, and the only peaks recorded from the gauze with any appreciable intensity are those for mass numbers 24, 25 and 26. Both spectra have small peaks at mass numbers 27 and approximately 43 from the cathode. It will be noticed that 26 from the cathode is less intense in the second spectrum, but it does not necessarily follow that this is due entirely to increase of V_1 , since it was noticed that with V_1 100 v. the peak height slowly decreased with time. It must be emphasized that it is not justifiable to make direct comparisons between the intensities of ion peaks

Table 1. Negative ions identified by different observers using oxide coated cathodes

1	2	3	4	5	6	7	8
1	—	—	H ⁻	H ⁻	H ⁻	H ⁻	H ⁻
12	C ⁻	C ⁻	C ⁻	C ⁻	—	C ⁻	C ⁻
13	—	—	CH ⁻	—	CH ⁻	—	—
14	—	—	—	—	CH ₂ ⁻	—	—
16	O ⁻	O ⁻	O ⁻	O ⁻	—	$\left\{ \begin{array}{c} \text{O}^- \\ \text{or} \\ \text{CH}_4^- \end{array} \right\}$	O ⁻
17	—	—	—	—	OH ⁻	—	—
18	—	—	—	—	H ₂ O ⁻	H ₂ O ⁻	—
19	—	F ⁻	F ⁻	F ⁻	—	F ⁻	—
23	—	—	—	—	Na ⁻	—	—
24	—	C ₂ ⁻	—	C ₂ ⁻	C ₂ ⁻	—	—
25	—	C ₂ H ⁻	—	C ₂ H ⁻	C ₂ H ⁻	—	—
26	CN ⁻	CN ⁻	CN ⁻	CN ⁻	C ₂ H ₂ ⁻	$\left\{ \begin{array}{c} \text{CN}^- \\ \text{or} \\ \text{C}_2\text{H}_2^- \end{array} \right\}$	CN ⁻ or C ₂ H ₂ ⁻
27	Al ⁻	Al ⁻	—	—	—	—	—
30	—	—	—	—	—	$\left\{ \begin{array}{c} \text{NO}^- \\ \text{or} \\ \text{HCHO}^- \end{array} \right\}$	—
32	O ₂ ⁻	O ₂ ⁻	O ₂ ⁻	O ₂ ⁻	O ₂ ⁻	O ₂ ⁻	O ₂ ⁻
33	—	—	—	—	?	—	—
34	—	—	—	—	?	—	—
35	³⁵ Cl ⁻	³⁵ Cl ⁻	³⁵ Cl ⁻	³⁵ Cl ⁻	³⁵ Cl ⁻	³⁵ Cl ⁻	³⁵ Cl ⁻
37	³⁷ Cl ⁻	³⁷ Cl ⁻	³⁷ Cl ⁻	³⁷ Cl ⁻	³⁷ Cl ⁻	³⁷ Cl ⁻	³⁷ Cl ⁻
39	—	—	—	? ³⁹ K ⁻	—	—	—
40	—	—	? Ca ⁻	—	Ca ⁻	Ca ⁻	—
42	CNO ⁻	—	—	—	?	—	$\left\{ \begin{array}{c} \text{CNO}^- \\ \text{or} \\ \text{C}_3\text{H}_6^- \end{array} \right\}$
43	?	—	—	—	?	$\left\{ \begin{array}{c} \text{diffuse} \\ 42.7^* \end{array} \right\}$	—
48	—	—	$\left\{ \begin{array}{c} \text{broad} \\ \text{diffuse} \\ \text{peak}^* \end{array} \right\}$	—	?	—	$\left\{ \begin{array}{c} \text{"suspicion"} \\ \text{of 48}^* \end{array} \right\}$
56	—	—	—	—	CaO ⁻	CaO ⁻	—
58	—	—	—	—	Ni ⁻	—	—
60	—	—	—	—	Ni ⁻	—	—
62	—	—	—	—	?	—	—
68	—	—	—	—	?	—	—
74	—	—	—	—	—	Ca(OH) ₂ ⁻	—
79	—	—	⁷⁹ Br ⁻	—	—	80 ?	$\left\{ \begin{array}{c} \text{unresolved} \\ \text{Br}^- \end{array} \right\}$
81	—	—	⁸⁰ Br ⁻	—	—	—	
101	—	—	—	—	—	$\left\{ \begin{array}{c} \text{CaCO}_3^- \\ \text{or} \\ \text{SrO}_2^- \end{array} \right\}$	—
120	—	—	—	—	—	SrO ₂ ⁻	—
127	—	—	—	—	—	—	I ⁻
168	—	—	—	—	—	BaO ₂ ⁻	—
261	—	—	—	—	—	Ba(NO ₃) ₂ ⁻	—

Cols. 2-5 : observations by Sloane and Watt.

2-3, using demountable system.

4-5, using glass sealed-off tubes.

2, 4, at cathode, 3, 5, at gauze.

Col. 1 : mass-to-charge ratio of ion.

Col. 6 : observations by Schaefer and Walcher.

Col. 7 : observations by Bachman and Carnahan.

Col. 8 : observations by Broadway and Pearce.

* See § 4.

for widely different mass numbers, since the spectra were plotted by varying the accelerating voltage. This means that different ion beams traversed the semi-circular arc to the Faraday cylinder with different energies and consequently possibly suffered different losses. In addition to this the ion optical effects in the source chamber were different. It was not convenient to arrange for good focusing without sacrificing the simplicity of design which appeared essential if the area of bombarded surfaces was to be kept small.

All negative ions detected by the use of the metal mass spectrometer have their identifications, in some cases tentative (see § 4), listed in columns 2 and 3 of table 1. Column 1 of the same table gives the mass-to-charge ratios (mass numbers) of these and other ions.

(ii) *Calibration of the mass scale for the metal spectrometer using a Kunsman positive ion source*

The value of M/e for a particular beam can be calculated roughly from the magnetic field, the accelerating voltage V_n and the geometry of the spectrometer; but to obtain a more definite calibration a special Faraday cylinder was used with a Kunsman alkali positive ion source (Tyndall 1938) fitted behind a small opening in its base. When the source was heated, alkali positive ions could be drawn through the opening and accelerated into the deflection chamber through the slits S_3 and S_2 . They could then be made to traverse the negative ion path in the reverse direction to S_1 , and be detected by measuring the current to the electrode system in the source chamber. This calibration could be effected, without altering the magnetic field, by merely interchanging the source supply leads and the electrometer lead. Any possible errors due to irreproducibility of the magnetic field were thus avoided.

Figure 4 shows the Na^+ calibration peak (a) together with the negative ion peaks (b) for mass numbers 26 from the cathode and 24, 25 and 26 from the gauze, for the particular case of V_1 equal to 100 v., and establishes definitely the M/e scale for the negative ions. This was confirmed by comparing the positions of the two potassium positive ion peaks (mass numbers 39 and 41) with those due to the two chlorine negative ions (mass numbers 35 and 37).

(iii) *Ion spectra : glass sealed-off tube*

A representative spectrum taken with the first glass sealed-off tube is shown in figure 5. In this case the cathode emission was 27 ma. and V_1 100 v. The gauze G_2 was connected to the slit S_1 . Using the same method for finding the place of origin of the ions as in the case of the demountable system, the ions listed in columns 4 and 5 of table 1 were identified. In the present case, the mass scale was fixed taking the chlorine isotope of mass 35 as standard, and ions of mass less than 12 were not studied. With V_1 120 v., other conditions being the same, the small peak apparently in region 32 from the cathode became resolved into two; one remained fixed and was therefore due to an ion of mass number 32 from the cathode, whilst the other which moved was due to an ion of mass number 39 from the gauze. It is for this reason that these peaks have been shaded as in figure 5. The diffuse unresolved peak in the region where V_n is approximately 370 v. may be due to ions from the cathode with mass numbers in the range 40 to 60. It does not move as V_1 is changed, but has not been shaded as it may include some ions from the gauze. The same applies to part of the unshaded

background for smaller masses. The diffuse peak will be discussed later (§4). No ions with mass number greater than 81 were observed although they were sought for using higher values of V_n and correspondingly higher values of magnetic field, with appropriately increased magnetic screening of the source chamber S.

Using a suitably small magnetic field, peaks due to hydrogen ions from the cathode and gauze appeared at the values of V_n shown in figure 6. For this spectrum V_1 was 120 v. and the cathode emission was 33.5 ma. These ions have

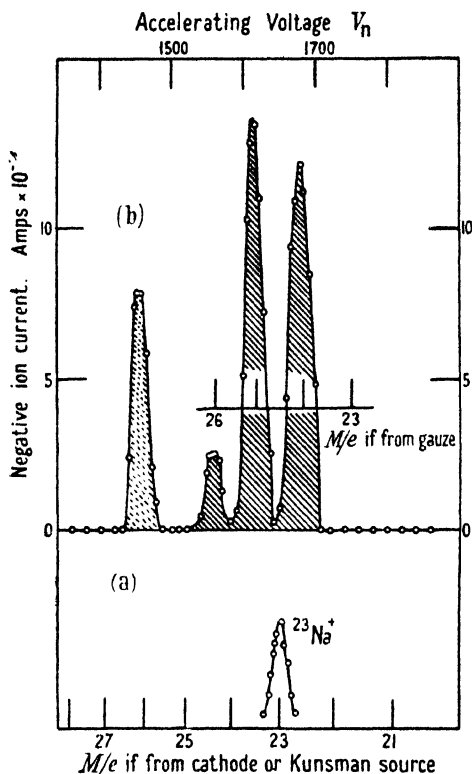


Figure 4. Calibration peaks for metal mass spectrometer. (a) Peak due to sodium positive ions from a Kunsman source placed behind the Faraday cylinder. (b) Peaks due to negative ions from the cathode and gauze, taken without altering the magnetic field in any way after taking (a). The shading of the peaks is the same as in figure 2. $V_1=100$ volts.

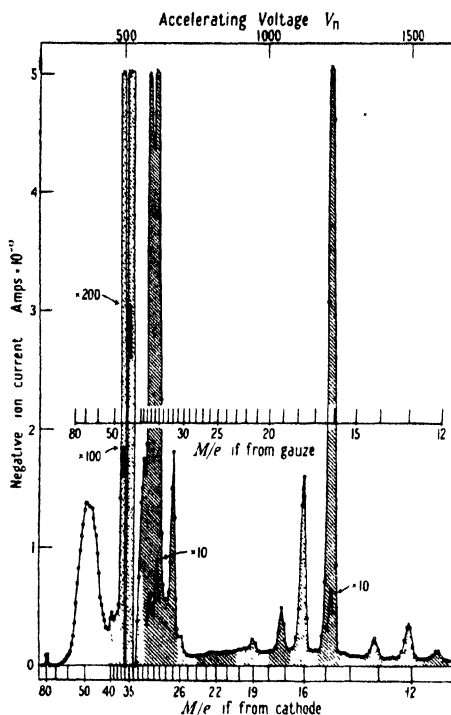


Figure 5. Negative ions mass spectrum taken with the first glass sealed-off tube shown in figure 1. The shading of the peaks is the same as in figure 2. $V_1=100$ volts.

been included in columns 4 and 5 of table 1. No ions with mass numbers between 12 and 1 were observed in this instance although it is known that $^6Li^-$ and $^7Li^-$ exist (Sloane and Love 1947).

After prolonged use of this tube a spectrum taken with voltages and magnetic field similar to those used for figure 5 revealed only the chlorine peaks from the cathode in any appreciable intensity, and even they had decreased until 37 showed a peak height of only 2.2×10^{-13} amp., compared with 1.85×10^{-11} in figure 5. The only other ions then detected were 16 from the cathode, 32, 35 and 37 from

the gauze, all less than 10^{-14} amp. peak height, and 16 from the gauze with a peak height of about 3×10^{-14} amp. Although the ion peak heights had decreased, the electron emission had increased for the same value of V_1 . The decrease in the heights of the peaks from the gauze may have been due either to the tube vacuum having become harder so that the number of residual positive ions available for bombarding the gauze was less, or to the gauze having become thoroughly degassed, since at one stage it was run red hot, or to a combination of these two effects.

(iv) Energy distribution

Arnot and Milligan (1936) found that a negative ion produced from a surface by bombardment with the corresponding positive ion could have an energy in excess of that imparted to it by the accelerating electric fields in the apparatus. Later Sloane and Press (1938), using their double mass spectrometer, showed that a given positive ion could liberate different negative ions and that, in the cases investigated, these also could have excess energies.

In the present work a retarding potential method (Arnot and Milligan 1936, Sloane and Press 1938) was used for studying the energy distribution in the various negative ion beams. Particles of a given kind differing in momentum, due to their leaving their place of origin with a range of initial velocities, could be brought on to the slit S_2 in front of the Faraday cylinder by varying the accelerating voltage V_n . The total number of ions of a given kind approaching the slits, which left the source (cathode or gauze) with more than any particular initial energy V_r , could then be found by holding the Faraday cylinder, and its screening slit S_3 , V_r volts negative to the source, whilst V_n was varied so that the whole inherently diffuse beam moved across the slits. Arnot and Milligan, using variation of magnetic field instead of V_n , found that the peak heights of the curves so obtained were proportional to the areas beneath the peaks, and it will again be assumed (Sloane and Press 1938) that this is also true when V_n is varied. The energy distribution curve of the ions is then obtained by first plotting a curve showing peak height against V_r , and then differentiating this with respect to V_r . Representative curves obtained in this way and showing the existence of initial energy, are shown in figures 7, 8, 9 and 10. Figures 7 and 8 were obtained using the demountable system and figures 9 and 10 using the first glass sealed-off tube. These results demonstrate clearly that a relatively large number of ions with mass numbers 16 from the cathode, 26 from the gauze and 1 from the cathode have considerable initial energy, while only a relatively small number of ions of mass number 1 from the gauze have any appreciable initial energy. The distribution curves for the ion of mass number 1 explain the different shapes of the two peaks in figure 6, the tail on the left hand side of the peak from the cathode being due to ions with initial energy. In the same way figure 7 explains the high energy tail to the left of 16 from the cathode in all spectra. Figure 11 shows a retarding potential curve, taken with the demountable

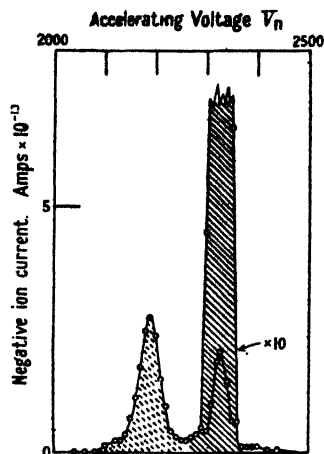


Figure 6. Peaks due to hydrogen atomic negative ions from the cathode (left) and gauze (right) in the first glass sealed-off tube. $V_1 = 120$ volts.

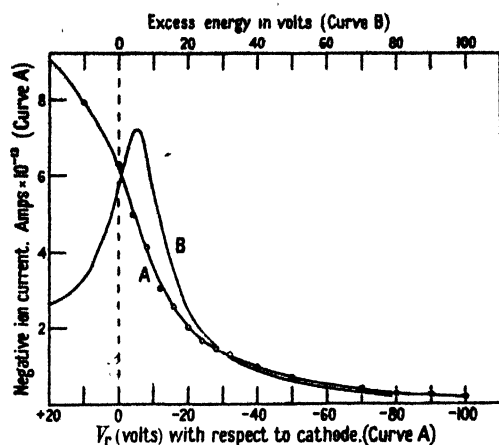


Figure 7. Retarding potential curve (A) and derived energy distribution curve (B) for oxygen ($M/e=16$) negative ions from the cathode in the metal demountable mass spectrometer. $V_1=200$ volts, cathode emission $=6.2$ ma., and $V_n=2232$ volts (connection to cathode) when V_r was zero.

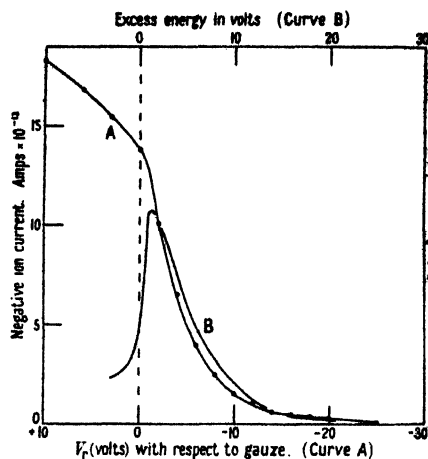


Figure 8. Retarding potential curve (A) and derived energy distribution curve (B) for negative ions with $M/e=26$ from the gauze in the metal demountable mass spectrometer. $V_1=200$ volts, cathode emission 10.3 ma., and $V_n=1447$ volts (connection to gauze) when V_r was zero.

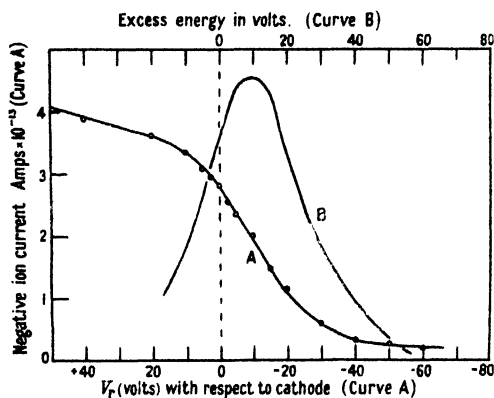


Figure 9. Retarding potential curve (A) and derived energy distribution curve (B) for hydrogen ($M/e=1$) negative ions from the cathode in the first glass sealed-off tube. $V_1=100$ volts, second gauze connected to slit, cathode emission 27.8 ma., heater current 1.3 A. (normal), and $V_n=2340$ volts (connection to cathode) when V_r was 60 volts accelerating.

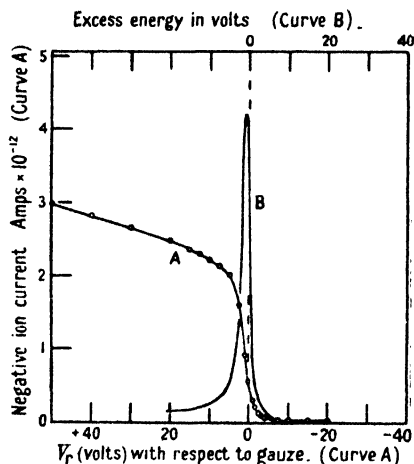


Figure 10. Retarding potential curve (A) and derived energy distribution curve (B) for hydrogen ($M/e=1$) negative ions from the gauze in the first glass sealed-off tube. Conditions are the same as for figure 9 except that $V_n=2365$ volts, the connection being made to the gauze.

system, for ions with no measurable initial energy, namely those of mass number 42 from the cathode. The curves for 35 and 37 from the cathode are similar except that they do not show the kink at the upper bend. The origin of the kink is unknown but it is not significant for the present discussion. All three curves fall abruptly when the Faraday cylinder is approximately 2.5 v. positive with

respect to the source, but if there were no contact potential differences and no fall of potential through the oxide coating they would presumably fall at zero volts. This stagger from zero gives a correction factor which should be applied when estimating the initial energies in the case of, for instance, figure 7, but it can frequently be ignored when the initial energies are of the order of 100 v. It should also be taken into account when setting the mass scales for ions from the gauze in all the mass spectra, but this refinement is unnecessary in the present instance.

When taking all these retarding potential measurements V_n was connected between S_1 and the source (cathode or gauze) of the ions of immediate interest, while V_r was applied by connecting a single variable small voltage between the source and the Faraday cylinder. Results obtained for all ions investigated are summarized in table 2.

Table 2. Summary of results for initial energy distribution

Ions from oxide cathode			Ions from tantalum gauze		
M/e	Apparatus	Initial energy	M/e	Apparatus	Initial energy
1	Sealed-off	Up to ≤ 60 v.	1	Sealed-off	Small tail up to ≤ 10 v.
16	Metal	Up to ≤ 100 v.	24	Metal	Up to ≤ 25 v.
26	Metal	Very small (< 4 v.)	25	Metal	Up to ≤ 8 v.
35	{ Metal	None	26	Metal	Up to ≤ 25 v.
	{ Sealed-off	None			
37	{ Metal	None			
	{ Sealed-off	None			
42	Metal	None			

Of the other ions from the cathode, those with mass numbers 12, 13, 19, 27, 40, 43, 79 and 81 gave peaks which were too small for satisfactory retarding potential experiments to be performed. The peak due to the ion of mass number 32 from the cathode was not investigated by the retarding potential method, but it can be seen from figure 2 (b) to possess a considerable excess energy tail. In the same way figure 2 shows the ion beams of mass numbers 16, 32, 35 and 37 from the gauze contain large numbers of ions with appreciable excess energies. This is confirmed by figure 5, which also indicates that there are some fast ions of mass numbers 12 and 19. The possession of excess energies by the chlorine isotopes from the gauze explains why the peaks due to these ions

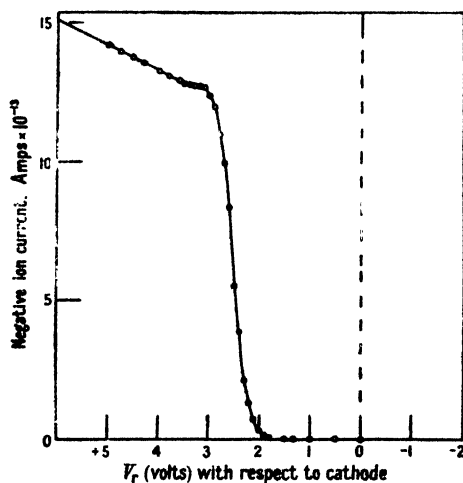


Figure 11. Retarding potential curve for negative ions with $M/e=42$ from the cathode in the metal demountable mass spectrometer. $V_1=200$ volts, cathode emission= 1.75 ma., and $V_n=850$ volts (connection to cathode). This exhibits no excess energies.

are not separated in figure 5, while those from the cathode, showing only an approximately Maxwellian distribution appropriate to the cathode temperature, are resolved.

(v) *Variation of chlorine peak heights with cathode temperature*

Figure 12 shows the rise and fall of the emission of $^{35}\text{Cl}^-$ and $^{37}\text{Cl}^-$ from the cathode for various cathode temperatures, determined by the heater currents shown. These curves were obtained using the first sealed-off tube after ageing, when the chlorine ions gave the only strong beams. V_1 was 150 v. The points shown are the maxima for each temperature (see next section). The curve relating electron emission and cathode temperature, which is also plotted in the figure, shows that when the heater current was raised above about 1.3 amp. the electron emission became space-charge-limited, although the ionic emission still increased rapidly. The subsidiary maximum and minimum appear to be real, since they repeat as shown. It may be significant that they occur at the points where the electron current becomes space-charge-limited, as they were absent from a similar set of curves taken with the second sealed-off tube under conditions when there was no space-charge limitation. The shoulders on the falling parts of the ionic curves in figure 12 were, however, present on the curves taken with the second tube.

(vi) *Time changes in emission after alterations in electric field or cathode temperature*

While investigating the negative ion mass spectra at different cathode temperatures and different values of V_1 , certain interesting time changes were noted. Using the first sealed-off tube under normal operating conditions for the cathode, when the cathode had aged, the variations of intensity of the only peaks observed from the cathode, those of chlorine, were investigated. It was found that, while the electron emission was space-charge-limited, on suddenly increasing the cathode temperature the emission of $^{35}\text{Cl}^-$ and $^{37}\text{Cl}^-$ increased immediately to a maximum and then fell *gradually* to a steady value, but on increasing V_1 the ion emission rose *gradually* to a maximum before falling slowly to an approximately constant value; the falling part of the curve often showed secondary maxima and minima. On the other hand, when V_1 was decreased from 100 to 50 v. the ion current fell immediately and quickly reached a steady value.

Similar time effects were observed with the slightly soft second sealed-off tube. Figure 13 shows the variation of the chlorine peak heights with time on changing V_1 in this tube. At X in this figure, V_1 was raised from 120 to 150 v.,

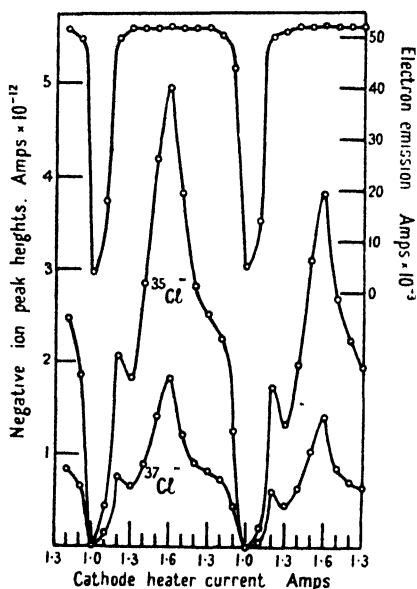


Figure 12. Variation of chlorine negative ion peak heights and electron emission (top curve) with cathode heater current, and hence cathode temperature. The heater current was lowered to 1.0 A. from its normal value of 1.3 A., raised to 1.6 A., reduced again to 1.0 A., raised again to 1.6 A., and finally reduced to 1.3 A. $V_1 = 150$ volts.

the electron emission increasing abruptly from 27.5 ma. to 39 ma. At Y, V_1 was raised to 180 v., the electron emission rising to 50 ma. At Z, V_1 was reduced again to 150 v., when the electron emission fell to 39 ma.

Somewhat similar results for the variation of negative ion emission with accelerating voltage have been recorded previously with a PbCl_2 source (Zentgraf 1929). It seems probable that the time changes in the emission of $^{35}\text{Cl}^-$ and $^{37}\text{Cl}^-$ from an oxide coated cathode are connected with electrolytic phenomena in the coating.

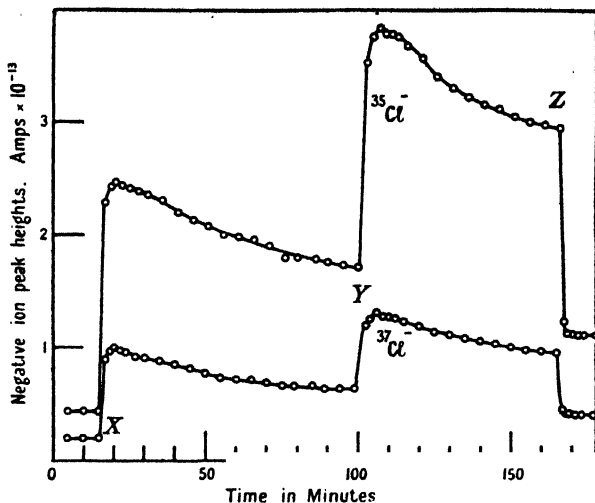


Figure 13. Variation of the chlorine negative ion peak heights with time after changes in V_1 in the slightly soft second sealed-off tube.

(vii) *Time changes in emission after change in cathode surface conditions due to positive ion bombardment*

The variation with surface conditions of the chlorine peak heights from the cathode and of the electron emission, for fixed values of V_1 and cathode temperature, could be studied by making use of the small amount of residual gas present in the second sealed-off tube. When preliminary measurements were being made with this tube it was observed that when the magnetic field was of such a value that large values of V_n were necessary to detect the chlorine ions, the electronic emission gradually decreased with time, while the peak heights of the chlorine ions gradually increased. This is believed to have been due to positive ions formed by electron collision in the residual gas bombarding the cathode and altering its surface. With small values of magnetic field and correspondingly small values of V_n this bombardment effect was negligible. These low values could therefore be used to observe the changes in emission due to recovery of the cathode after it had been bombarded for a long period using high values of V_n .

Repeated observations showed that on reducing V_n from a large to a small value suddenly, leaving V_1 and the cathode temperature unaltered, the chlorine peak heights decreased gradually with time while the electronic emission gradually increased.

§ 4. DISCUSSION

The ions detected by Bachman and Carnahan (1938), Broadway and Pearce (1939), Schaefer and Walcher (1943) and by us are listed for comparison in table 1. Column 1 of table 1 gives the mass-to-charge ratios (mass numbers) of all ions detected in any of the above investigations; columns 2 and 3 give the ions found by us to originate at the cathode and gauze respectively in the metal apparatus, while columns 4 and 5 list similar results for our sealed-off tubes; columns 6, 7 and 8 give the ions identified by Schaefer and Walcher, Bachman and Carnahan, and Broadway and Pearce respectively. It has been

assumed that negative ions with more than one electronic charge are unstable (Massey 1938). The identification of the ions with mass numbers 1, 12, 16, 19, 32, 35, 37, 79 and 81 appears to be definite, either from their isolation in the spectra (1, 12), their known electronegative character (16, 19), the isotope ratios (35, 37, 79, 81), or from a combination of these. The peak of mass number 32 has usually been ascribed to O_2^- , an ion which has been considered in detail (Bates and Massey 1943), but, as mentioned by Broadway and Pearce, it might be due in part to S^- , and sulphur is still believed to form a stable negative ion (Bates 1947). The identification of most of the others must be more tentative, although a good case could be made for the persistent ion at 26 being CN^- , on account of the electronegative properties of this radicle. $^{27}Al^-$ has been reported previously from an aluminium oxide source (Blewett and Jones 1936), and it is thought that this ion is stable (Bates 1947), but it is not certain that in these spectra the peak at 27 is not due to the ion of some hydrocarbon or other radicle. Schaefer and Walcher's identification of Na^- requires further investigation (Sloane and Love 1947). The same applies to the ion of mass number 39 which has been found using the first sealed-off tube, and has been reported previously by Ditchburn and Arnot (1929) and attributed tentatively to $^{39}K^-$. The identification of the organic negative ions is necessarily very tentative. The two principal ion groups studied with low resolving power by Sloane and Press (1938), where exact identification was of secondary importance, and interest was centred on the process of formation, are seen by comparison with the spectra described in this paper to have been almost certainly composed of ions with mass numbers 24, 25 and 26 in the one case, and 35 and 37 in the other.

Bachman and Carnahan mention that, in the case of a tube sealed-off after breakdown of the cathode, only ions of mass numbers 16, 26, 35 and 37 were observed, while 16 and 26 gradually disappeared. This is in accord with our experience while using the first sealed-off tube when, after considerable use, only 35 and 37 from the cathode could be detected with any appreciable intensity under normal conditions, and it is to be contrasted with the statement by Schaefer and Walcher that after a few hours 35 and 37 disappeared and were undetectable in their tube, although other ions persisted.

Bachman and Carnahan recorded a spot at mass number 42.7 which Bachman considered (private communication) might be complex. Our figure 3, obtained using the metal mass spectrometer, shows, in fact, this region resolved into 42 and 43, although in our apparatus 42 was much more intense than 43. Again, Bachman and Carnahan mention that this spot sometimes became smeared out into a large area. Figure 5 shows that we observed a diffuse peak in the region of 50 before the first sealed-off tube became aged, and a similar peak appeared with the second sealed-off tube. It is possible that this diffuse peak is to be correlated with the smeared-out spot referred to by Bachman and Carnahan, and a spot suspected by Broadway and Pearce at 48. Its origin is not clear.

Some of the spots observed by Bachman and Carnahan and by Broadway and Pearce were sharply focused while others were not. Broadway and Pearce concluded that the sharply focused spots were due to negative ions emitted directly from the cathode, whereas Bachman and Carnahan considered that they were due to ions arising on the cathode surface close to the axis of the electron gun as a result of positive ion bombardment. Our results show that some ions from the cathode, particularly those with mass numbers 35, 37 and 42, show no

excess energy and have apparently an approximately Maxwellian distribution of energy appropriate to the cathode temperature, while 26 shows a diminutive excess energy tail extending only to 4 v. These ions would therefore be expected to give sharply focused spots in cathode-ray tubes with electrostatic focusing; in fact, Bachman and Carnahan remark that 26, 35 and 37 were sharply focused, although Broadway and Pearce found 26 and 42 slightly diffuse. Further, these energy distributions support the view that 35 and 37 are emitted thermionically as a result of electrolytic action in the cathode. Additional support is afforded by our observations on the variations of the heights of the peaks due to the chlorine ions when either V_1 or the cathode temperature is altered (§3(vi)). The ion with mass number 42 may also be emitted by this process and 26 may be mainly so emitted.

There is, however, another possible mode of formation for these two ions. It is known (Broadway and Pearce 1939) that oxide cathodes readily adsorb particles of an electronegative character and, although the surface layer so formed appears to be very stable, some re-evaporation presumably takes place; it is possible that some of the released particles are negatively charged. The CNO^- ions, presumed to be responsible for the peak at mass number 42, are perhaps released by this mechanism. They appeared in our work only in the metal mass spectrometer, which was sealed with grease (unless indeed the broad diffuse peak at 50 in figure 5 is at least partly due to them) and this agrees with the observation by Broadway and Pearce that these ions appeared most strongly in the presence of grease vapours. The CN^- ions with mass number 26 may also be emitted mainly by this re-evaporation process, but the diminutive excess energy tail on their energy distribution curve is probably due to ions liberated by positive ion bombardment, unless it just possibly results from the dissociation of a larger molecule. There does not, however, appear to be any intimate connection between the intensities of 42 and 26 either in the present work or in that of Broadway and Pearce. This slight excess energy would account for Broadway and Pearce's observation of a slightly diffuse spot due to ions with mass number 26, if the bombardment process was enhanced in their tube in the case of this ion.

Both Bachman and Carnahan, and Broadway and Pearce find the O^- spots diffuse. Bachman and Carnahan explain this as being due to a large area of the cathode being bombarded, while Broadway and Pearce think it may be due to the ions concerned being formed by dissociation of O_2 after capture of an electron, or by conversion of oxygen positive ions (Arnot and Milligan 1936). Our retarding potential results for O^- show that this ion from the cathode has a large excess energy tail on the energy distribution curve, with some ions possessing up to 100 v. excess energy. It therefore seems probable that the diffuseness of the spots is due to this large energy spread, since sharply focused spots can result only from beams homogeneous in energy. The H^- ions from the cathode have a somewhat similar energy distribution, and this almost certainly accounts for the diffuseness of the H^- spot also referred to by Broadway and Pearce. We did not examine the peak at mass number 32 by the retarding potential method, but spectrum (b) of figure 2 shows this peak to have the curved low voltage side characteristic of peaks including ions with excess energy. Bachman and Carnahan state that this spot is diffuse. They also say the spot due to 12 is diffuse, but in our case the peak due to 12 was too faint for energy distribution.

runs although its shape in the mass spectrum is consistent with its resulting from a beam including some ions with excess energy. Ions liberated from surfaces by positive ion bombardment seem always to be characterized by the possession of large excess energies (Arnot and Milligan 1936, Sloane and Press 1938), and, if this is so, it seems reasonable to conclude that at least a large part of each of 32, 16, 12 and 1 from the cathode is due to this process.

The observation by Becker (1929) and Detels (1927) that most of the oxygen emitted when space current is drawn is uncharged, although some ions are emitted, is consistent with the above conclusion if we envisage the oxygen flux as maintaining a surface layer which is then "sputtered" off (Sloane and Press 1938) as negative ions by the positive ion bombardment.

The oxygen peaks, among others, disappeared as our first sealed-off tube became aged, but when V_1 was increased and a large electron current was drawn from the aged cathode, 16 reappeared strongly. This result is in agreement with Becker's (1929) observation that oxygen can only be released from a well aged cathode by drawing a strong electron current. Schaefer and Walcher (1943) recorded O_2^- but not O^- . Barton (1925) also reported this phenomenon in the case of well formed and aged cathodes. However, his resolution was not high and our experience with the first sealed-off tube when aged gives experimental support to the suggestion by Broadway and Pearce that the ion observed by Barton may in fact have been chlorine. Blewett and Jones (1936) also detected O_2^- from a barium oxide emitter, but under conditions when chlorine contamination was certainly present.

Bachman and Carnahan (1938) mentioned that some spots were surrounded by small circles, in particular those due to mass numbers 26, 35 and 37, and that the circles might be "evidence furthering the theory of a gas source other than the cathode". Later Sloane and Cathcart (1939) pointed out that these circles might be due to bombardment of a gun aperture by positive ions and just possibly by negative ions from the cathode. The latter process was suspected by Sloane and Cathcart because in their early work the absence of ions with excess energy in the 26, 35 and 37 beams from the cathode was taken to show that the cathode was not bombarded by positive ions. In the apparent absence of positive ion bombardment Sloane and Cathcart were driven to the conclusion that the ions detected from the gauze were liberated under negative ion bombardment. Shortly afterwards however, they detected the excess energy tail on 16 from the cathode so that it appeared probable that positive ion bombardment played an important, if not the only, rôle. In a subsequent paper, Bachman (1940) pointed out that the bombarded aperture was the so-called grid and that it was at a potential negative to the cathode, so that negative ions would pass it by in favour of the first anode. He then proceeded to give a detailed explanation of the formation of the observed ring structure. In particular, he attributed the diffuseness of the rings to bombardment of the upper and inner sides of the aperture, in addition to the sharp aperture edge, by unfocused positive ions formed in the gas in the region between the grid and the first anode. Since the intensity distributions on proceeding radially through the ring patterns on the screen varied widely from ring to ring in a given spectrum it appears to us that the diffuseness is more likely to be due to the excess energies observed by us for all ions from the gauze which could be examined, as we have found that the energy distribution curves vary markedly from ion to ion. Bachman

also found that some rings showed a focused spot in the radially outermost section, while others did not. These spots he explained as being due to negative ions liberated from the aperture as a result of bombardment by a focused positive ion beam originating in the gas in the region between the first and second anodes. The appearance of these spots on some rings and not on others he explained by assuming that certain positive ions, present in the unfocused grid-first anode swarm responsible for the ring, were absent from the focused first anode-second anode beam. This explanation if correct makes his observations very interesting in that, while it is certain that bombardment of a contaminated surface by a given positive ion can give a whole spectrum of negative ions (Sloane and Press 1938), Bachman's observations support the view that this spectrum depends to some extent on the type of positive ion bombarding the surface.

Schaefer and Walcher (1943) list ions of mass numbers 24 and 25 from their oxide coated cathode, but since their cathode and the neighbouring grid were at the same potential they could not distinguish between these two electrodes as possible sources of ions. Our results suggest that 24 and 25 detected by them really arose from positive ion bombardment of the grid. We have never observed 24 or 25 from an oxide coated cathode, in agreement with Bachman (private communication). Schaefer and Walcher state that all peaks were sharp and mention that an energy spread of ± 10 v. at M/e 25 would have produced noticeable broadening. This is not compatible with our results.

Figure 12 shows that as the cathode temperature is increased, while V_1 is held constant, the chlorine ion current continues to increase rapidly with cathode temperature even after the electronic emission has become space-charge-limited. Schaefer and Walcher (1943) found a similar effect while investigating the dependence on cathode temperature of the emission of the ion ascribed to O_2^- , in the case of a formed cathode. The curves in figure 7(b) of their paper show exactly the same trend as do those of our figure 12 when the heater current, and hence the cathode temperature, is increasing. It will be noticed that in our case, although the electron emission reaches the same values on the repeated cycle of heating and cooling, the chlorine ionic emission appears to be falling progressively. These facts might be interpreted as indicating that the electron and ionic emissions are not intimately connected, but Liebmman (1946 a, b) has shown that the electron optical emission picture of the thermionic cathode on the cathode-ray screen is reproduced in negative quite faithfully, except for a few details, by the ion burn; a fact which he apparently rightly interprets as indicating that the *local* electron current density at the cathode surface and the rate of negative ion production are in most cases closely correlated.

The results described in §3(vii) show that bombardment of the cathode by positive ions reduces the electronic emission, which subsequently recovers when the bombardment ceases. This effect was also observed by Schaefer and Walcher, using hydrocarbon positive ions for bombardment. They also found that the CH^- ion beam, the only one investigated by them in this connection, was not increased by the bombardment, but our results show definitely that bombardment increases the intensity of the chlorine ion beams, and that these decrease again when the bombardment is discontinued.

Throughout most of this discussion it has been assumed that the surface production of negative ions is due to one of two causes; either they are evaporated thermionically or they are sputtered off by bombardment with positive ions.

In the former case the energy distribution of the emitted ions will be thermal with a temperature that of the surface. In the latter case the energies are distributed over a very much wider range, the exact dependence of which on the masses of the bombarding and emitted ions, and on the energy of the bombarding ions, will be discussed in detail elsewhere. With both types of emission the question as to how the surface film is formed has been left practically open in this work. This question involves two distinct problems: the one that of migration of material to a surface from the interior of a solid, and the other that of the conditions of deposition of a surface film from low pressure gas. Both lie outside the scope of this investigation, but whatever mechanisms are ultimately found to explain these will require to be in accord with the results which have been obtained by us. It has also to be remembered that the study of negative ions as a whole, unlike that of positive ions, has still progressed comparatively little (Massey 1938, Bates and Massey 1943, Bates 1947), and that it may be found that processes other than these two may play a not unimportant part, as is in fact implied in Broadway and Pearce's discussion.

ACKNOWLEDGMENTS

The authors wish to thank Dr. Eliza B. Cathcart for assistance while taking observations during the early part of this work. They also wish to thank Professor K. G. Emeléus for continued interest, and are greatly indebted to the late Sir Clifford Paterson, F.R.S., Director of the Research Laboratories of the General Electric Company Ltd., for kindly providing the oxide coated cathodes used in all the experiments.

REFERENCES

- VON ARDENNE, M., 1935, *Arch. Electrotech.*, **29**, 731.
 ARNOT, F. L., 1937 a, *Proc. Roy. Soc. A*, **158**, 137; 1937 b, *Ibid.*, **158**, 157.
 ARNOT, F. L., and BECKETT, C., 1938 a, *Nature, Lond.*, **141**, 1011; 1938 b, *Proc. Roy. Soc. A*, **168**, 103.
 ARNOT, F. L., and MILLIGAN, J. C., 1936, *Proc. Roy. Soc. A*, **156**, 538.
 BACHMAN, C. H., 1940, *J. Appl. Phys.*, **11**, 83; 1945, *Gen. Elect. Rev.*, **48**, 13.
 BACHMAN, C. H., and CARNAHAN, C. W., 1938, *Proc. Inst. Radio Engrs.*, **26**, 529.
 BARTON, H. A., 1925, *Phys. Rev.*, **26**, 360.
 BATES, D. R., 1947, *Proc. Roy. Irish Acad.*, **51**, 151.
 BATES, D. R., and MASSEY, H. S. W., 1943, *Philos. Trans. Roy. Soc.*, **239**, 269.
 BECKER, J. D., 1929, *Phys. Rev.*, **34**, 1323.
 BLEWETT, J. P., and JONES, E. J., 1936, *Phys. Rev.*, **50**, 464.
 BROADWAY, L. F., and PEARCE, A. F., 1939, *Proc. Phys. Soc.*, **51**, 335.
 DETELS, F., 1927, *Jb. drahtl. Telegr.*, **30**, 10.
 DITCHBURN, R. W., and ARNOT, F. L., 1929, *Proc. Roy. Soc. A*, **123**, 516.
 LEVY, L., and WEST, D. W., 1936, *J. Instn. Elect. Engrs.*, **79**, 11.
 LIEBMANN, G., 1946 a, *Nature, Lond.*, **157**, 228; 1946 b, *Electronic Engng.*, **18**, 289.
 MASSEY, H. S. W., 1938, *Negative Ions* (Cambridge: University Press).
 SCHAEFER, H., and WALCHER, W., 1943, *Z. Phys.*, **121**, 679.
 SHARPE, J., 1946, *Electronic Engng.*, **18**, 385.
 SLOANE, R. H., and CATHCART, E. B., 1939, *Nature, Lond.*, **143**, 474.
 SLOANE, R. H., and LOVE, H. M., 1947, *Nature, Lond.*, **159**, 302.
 SLOANE, R. H., and PRESS, R., 1938 a, *Nature, Lond.*, **141**, 872; 1938 b, *Proc. Roy. Soc. A*, **168**, 284.
 TYNDALL, A. M., 1938, *Mobility of Positive Ions in Gases* (Cambridge Phys. Tracts).
 ZENTGRAF, M., 1929, *Ann. Phys., Lpz.*, **2**, 313.

Some New Aspects of High-Speed Oscillography

By D. M. MACKAY

Wheatstone Physics Laboratory, University of London, King's College

ABSTRACT. From a consideration of the mechanism of the cathode-ray oscillograph at very high speeds it is shown, in Part I, that conventional deflection systems do not fully realize the possibilities of the cathode-ray tube.

By a modification of the shape of the deflector-plates it is theoretically possible, under suitable conditions, to expand the time-scale obtainable with a given time-base generator up to infinity. This gain is obtained at the expense of definition, but brightness is not affected. Relativistic considerations are shown to be irrelevant.

The geometry of the deflected electron beam is examined, and an expression derived relating the shape of the effective time-base voltage wave-form, the time-scale, and the shape of the Y-plates.

Part II examines some practical factors ignored in the analysis of Part I. These cause various forms of trace distortion, which are classified as (i) intermodulation effects, (ii) mismatching distortion, (iii) transit-time effects, (iv) defocusing and equivalent effects.

Possible ways of overcoming a number of these difficulties are described. Some have already been employed with success, but others have not yet been followed up.

PART I: THE IDEAL CASE

§1. INTRODUCTION

THE upper limit to the attainable speed of a cathode-ray oscillograph is normally set by a combination of practical factors such as trace brightness, and theoretical limitations fundamentally inherent and insurmountable. It is characteristic of modern physics that the distinction between these two categories is rapidly disappearing, as analysis reveals the interrelation of their ultimate causes.

In the field of high-speed oscillography, however, it is still possible to draw a fairly well defined dividing line, and to consider separately the ideal case and the practical factors modifying it. The object of Part I is to examine the mechanism of the ideal cathode-ray oscillograph at high speeds, and to suggest a possible method of increasing the time-scale factor, theoretically up to infinity, without a corresponding increase in time-base voltage-gradient. Contrary to popular belief, it is shown that relativistic considerations set no limitation on writing speed, and that at high speeds the latter need not have any connection with the time-scale factor.

In Part II it is intended to examine some of the practical limitations of the technique, and to suggest some methods of at least partially overcoming them.

§2. THE NATURE OF THE DEFLECTED BEAM

Figure 1 represents part of a typical electrostatic deflection system in an "ideal" C.R. tube. A stream of electrons, assumed to have negligible cross-sectional area, moves in the direction OA with velocity v_f , and passes through a deflecting field E between two plates X_1 , X_2 , of negligible axial dimensions.

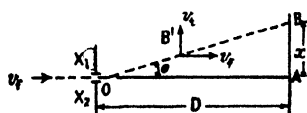


Figure 1.

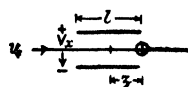


Figure 1 (a).

The electrons receive a transverse velocity v_t due to E , and travel at an angle θ to OA to strike a plane screen AB , normal to OA , at a point B whose distance from the axis is x' .

It is evident that if the distance OA be D ,

$$x'/D = \tan \theta = v_t/v_f. \quad \dots\dots(1)$$

Thus if v_f is constant,

(a) *The transverse velocity v_t of an electron producing a spot of light at a distance x' from the origin A is determined only by that distance, i.e. in other words, the maximum electron velocity which has to be attained under given conditions of gun voltage etc. is a function only of trace length, and is independent of writing speed.*

Differentiating equation (1), we have $dx'/dt = (D/v_f)(dv_t/dt)$, so that, under the same conditions,

(b) *The speed of the spot of light is determined by the rate of change of the transverse velocity imparted to successive electrons by the deflecting field.* This means that relativistic considerations do not affect the problem any more at high writing speeds than they do in the case of a steady deflection of the spot; the limit to writing speed, in the conventional oscillograph,* is set only by the available transverse acceleration.

What happens to the electron beam, then, when the writing speed approaches that of light? Evidently if the time taken for the spot to traverse the screen is short compared with the transit time from plates to screen, the stream of electrons producing the trace AB must form an expanding "kink" (as shown at $A'B'$ in figure 2), travelling towards the screen with an axial velocity v_f . When the leading edge of this kink reaches the screen, a spot of light will be formed at the point of intersection, which will travel rapidly along the line AB with velocity $v_s = v_f \cot \phi + v_t$, or $v_s \simeq v_f \cot \phi$, where ϕ is the angle between screen and kink at the point of intersection (figure 2(a)).

This physical picture suggests an interesting possibility: if the screen, instead of being normal to the tube axis, were inclined to it in such a way as to reduce ϕ , then v_s could be increased to any desired value up to infinity, without any increase in the transverse acceleration of the electron stream.

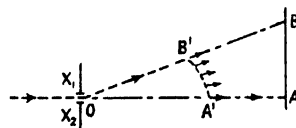


Figure 2.

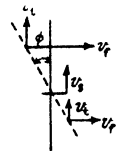


Figure 2 (a).

That such an arrangement could give an infinite writing speed is undeniable. But unfortunately—apart from the geometrical distortion of the picture which would occur at any speed—the tilting of the screen, and the increased writing speed caused by it, would have no visible effect on the picture, least of all on its "time-scale factor" (§3), for the spatial configuration of the electron stream is determined once the electrons have left the deflecting system, and the only effect of inclining the screen is to alter the speed of *delineation* of a picture whose time-scale is already fixed.

Evidently, in connection with high-speed oscillography, the electron stream can no longer be regarded as a beam or pointer pivoted at the deflector system, but rather as a formation of electrons setting out to delineate a picture on the screen or photographic plate. The course of each electron is predetermined by the deflector system, and the instant at which it strikes the screen is relatively

* Throughout this paper the term "oscillograph" will be used for convenience to cover both visual and photographic delineators of wave-form.

unimportant. The electrons can even arrive in reverse order, so that the spot of light travels in the negative direction of the time-scale (if ϕ is negative), without in any way affecting the appearance of the trace.

§ 3. THE TIME-SCALE FACTOR

The classification of the numerous terms used to describe the "speed" of an oscillograph—time-base speed, scan speed, sweep speed, writing speed, trace speed etc.—is by no means easy, since their connotation varies with different authors and different contexts. It seems clear, however, that there are already too many terms, and it is therefore with some trepidation that one ventures to suggest yet another, to represent a quantity which is not unambiguously defined by any of the above, but which is a parameter of fundamental interest to the user of an oscillograph.

An oscillograph with electrostatic deflection may be regarded as a device with two terminals between which a voltage $v_{in} = f(t)$ is applied (figure 3). It

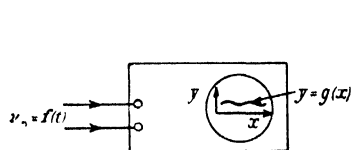


Figure 3.

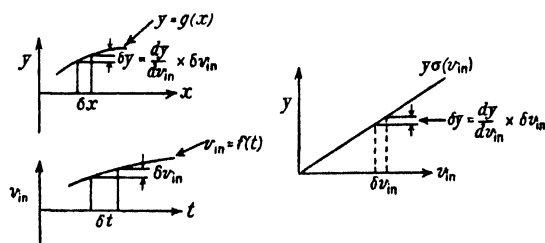


Figure 4.

produces on a screen a graph with the equation $y = g(x)$, where y is a function $\sigma(v_{in})$ of v_{in} , generally of the form

$$y = av_{in} + b. \quad \dots\dots(2)$$

The fundamental criterion of speed, which we shall refer to as the *time-scale factor* S , may be defined formally as the ratio of the interval δx between two adjacent ordinates y and $(y + \delta y)$ on the screen, to the interval δt between the two corresponding values of the input voltage, v_{in} and $(v_{in} + \delta v_{in})$ (figure 4). In the limit, we may write

$$S = (dx/dy)(dy/dv_{in})(dv_{in}/dt). \quad \dots\dots(3)$$

This quantity is the factor by which the X-scale must be multiplied to give the time-scale; it is that normally implied when one speaks of an oscillograph as having a writing speed of so many centimetres per microsecond:

It should be noted that S , as defined, has no relation to the speed of the spot delineating the wave-form. S is obtained theoretically by comparing corresponding intervals dx and dt on the two graphs, but dx/dt cannot be measured directly, since x as defined is not an explicit function of t , and is *not* the abscissa of the delineating spot. Accordingly equation (3) defines S in terms of three pairs of related quantities, the slopes of the x - y , y - v_{in} and v_{in} - t graphs, which can be measured. Writing $dy/dx = g'(x)$, $dy/dv_{in} = \sigma'(v_{in})$, $dv_{in}/dt = f'(t)$, we have

$$S = \sigma'(v_{in})f'(t)/g'(x). \quad \dots\dots(4)$$

If equation (2) holds, $\sigma' = a$, and $S = af'(t)/g'(x)$.

In conventional low-speed oscillographs it is unnecessary to distinguish between this time-scale factor and the X-component of the writing speed, but in high-speed oscillographs the two quantities may differ considerably in magnitude, and even in sign (see also § 5).

§ 4. Y-DEFLECTION

We have seen that an increased writing-speed need not imply an expanded time-scale. It will now be shown that at high speeds a simple relation exists between the shape of the Y-deflection system and the time-scale factor, as a result of which, conversely, a great expansion of time-scale may be obtained without any corresponding increase in writing speed (Mackay 1946 a).

The simplified arrangement shown in figures 1 and 2 does not include any means of deflecting the electron stream in the Y-direction—i.e. out of the plane of the diagram. In this paper we shall consider only electrostatic deflection systems for this purpose, and figure 5 illustrates the conventional arrangement,

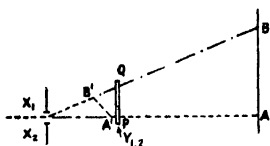


Figure 5.

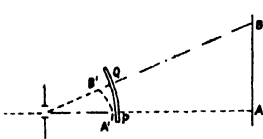


Figure 6.

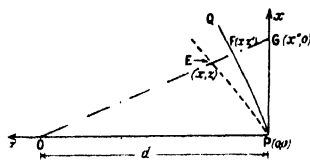


Figure 7.

in which a second pair of plates, Y_1 and Y_2 , is used to produce a deflecting field at right angles to the tube axis and to the direction of the X-deflection. The axial dimension of the Y-plates is assumed negligibly small, so that the Y-deflection of a given electron may be regarded as taking place when that electron traverses the line PQ representing the locus of the deflecting field. For convenience the trace is assumed to start at the centre of the screen.

In the case of high-speed deflection considered (figure 5), the electron at A' which is due to produce a spot of light at A receives its Y-deflection at the point P where $A'A$ intersects the Y-plates; and the electron at B' , similarly, receives its deflection at Q . *The interval of time portrayed on the screen between A and B is therefore the interval between the instants of intersection of electrons A' and B' with the deflecting field.**

In other words, the important quantity is the transverse speed of the point of intersection of the electron stream with the line PQ.

If now we apply the "inclined approach" technique to the Y-plates, we have in this case the theoretical possibility of attaining an infinite time-scale factor. All that is necessary is to reduce the interval of time between the instants of intersection of A' and B' with the deflecting field. This can be done by bending the line PQ towards $A'B'$, as shown in figure 6, so that Q comes to meet B' , and the velocity of the point of intersection of $A'B'$ with PQ is increased in the same way as the writing speed in the previous case, *and without any corresponding decrease in trace brightness.*

We have thus increased S (equation (4)) by reducing $g'(x)$, since $g'(x)$ tends to zero as the shape of the kink approaches that for which all electrons receive the same Y-deflection. Alternatively, one may say that the time-scale of the phenomenon, as "seen" by the electron stream, has been contracted.

* Subject to the conditions of § 5.

§ 5. THE EXPANDED TIME-SCALE FACTOR

To derive the relation holding between the shape of the plates, the time-base voltage wave-form, and the resulting time-scale factor S , it is convenient to refer the X-deflection of the spot to a plane normal to the Y-axis and passing through P (figure 7), and to take P as origin of an (x, z) coordinate system with the positive direction of the z -axis to the left of P.

The incident electron stream moving from left to right with a constant z -axial component of velocity v_f is represented by EP, where E (x, z) is a typical electron, whose path intersects PQ at F (x', z') and the x -axis at G (x'', O) . All electrons are assumed to receive x -deflection at the point O on the z -axis, the distance OP being d .

The time interval represented by the distance PG (x'') , or its projection on the screen, will be the time taken by an electron to travel from E to F, less the time taken for the signal to travel along the Y-plates from P to Q (or the difference in times of arrival of the signal at P and Q). For any given value of v_f and arrangement of Y-plates, the latter time correction can be allowed for as a z -correction in the equation for PQ. We shall henceforth assume that the equation of PQ,

$$z' = \Psi(x') \quad \dots\dots(5)$$

represents this "effective locus" of the Y-deflecting field.

With this assumption, we can state that the time interval T corresponding to PG is given by $T = (z - z')/v_f$, so that the time-scale factor

$$S = dx''/dT = v_f dx''/d(z - z'). \quad \dots\dots(6)$$

The writing speed, on the other hand, is simply $v_f dx''/dz$.

The time-base (X-deflecting) field at O has been responsible for imparting different transverse velocities v_x to successive electrons passing through it, and each value of v_x thus fixed determines a corresponding value of x'' , since $x''/d = v_x/v_f$. We can therefore express the action of this field in terms of x'' and time, or (since the time t between the passage of the electron at P and that at E through the X-field is $t = z/v_f$) in terms of x'' and z .

Thus instead of defining the X-deflection by the equation $v_x = V(t)$, we may write $x''v_f/d = V(z/v_f)$ or, with z in terms of x'' ,

$$z = v_f V^{-1}(x''v_f/d), \quad \dots\dots(7)$$

where V^{-1} is the reciprocal operator to V .

We shall find it convenient to rewrite equation (7) as

$$z/v_f = U(x''). \quad \dots\dots(8)$$

The equation of PQ, (5), may also be written in terms of x'' , for

$$x' = x''(1 - z'/d), \quad \dots\dots(9)$$

so that $z' = \Psi[x''(1 - z'/d)]$, which we may write as

$$z'/v_f = \Phi(x''). \quad \dots\dots(10)$$

Hence the interval T corresponding to x'' is given by

$$T(x'') = U(x'') - \Phi(x''). \quad \dots\dots(11)$$

This simple relation makes it possible, for a given plate shape, to derive the graph of x'' against T (the slope of which is S) merely by subtracting two ordinates,

provided that the X-deflecting wave-form is known. Alternatively the shape of plates required to give any characteristic can be derived directly from the given U and T curves (figure 8).

Equation (5) can conveniently be expressed in terms of the above functions, since

$$\frac{1}{S} = \frac{dT}{dx''} = \frac{1}{v_f} \frac{dz}{dx''} - \frac{1}{v_f} \frac{dz'}{dx''}$$

or

$$1/S = T' = U' - \Phi' *$$

A particular case of some interest arises when a linear time-base generator is required to produce a linear time-scale, expanded by the technique described.

In this case $T(x'')$ and $U(x'')$ are both proportional to x'' , so that from (11), $\Phi(x'') = U - T = kx''$. Now $z' = v_f \Phi = v_f kx''$ (from (10)), or from (9),

$$z'(d - z') = dv_f kx';$$

PQ therefore must form an arc of a parabola symmetrical about the line $z = d/2$.

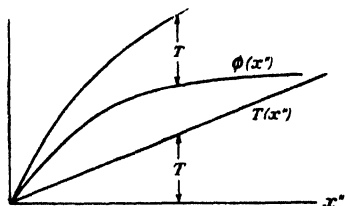


Figure 8.

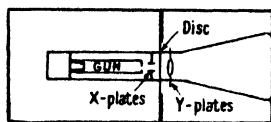


Figure 9.

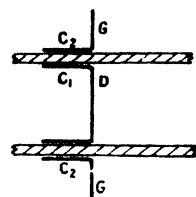


Figure 10.

The fact that Y-deflection does not take place at a constant distance from the screen implies that the Y-sensitivity will tend to vary as a function of X-deflection. This can be corrected for in the design of the plates, and should be borne in mind in relation to the second part of this paper.

§ 6. CONCLUSION

In this Part we have been concerned only with theoretical consideration of the ideal case in which many simplifying assumptions have been made. It is intended in Part II to discuss some of the practical limitations of this method of time-scale expansion. It is important, however, to note that even in the ideal case the technique is useful only when the time of rise of the deflecting voltage is already fairly short, since the complete kink A'B' must occupy only a fraction of the distance between plates and screen.

Assuming a value for v_f of 3×10^9 cm/sec., this means that if the kink is to occupy not more than, say, 10 cm. in the axial direction, it must be formed in less than $10/(3 \times 10^9)$ sec., or about $1/300 \mu$ sec. Thus the technique requires initially a really high-speed time-base generator, such as a hydrogen thyatron circuit, in order to be effective.

The author has obtained a time-scale factor exceeding the speed of light with the aid of such a circuit, and a normal deflector system. In this case the kink must have been formed in less than $1/1000 \mu$ sec., so that a time of $1/300 \mu$ sec. is evidently quite attainable with present-day components. Once it is attained

* This equation deals with reciprocals of speed, for which no term appears as yet to have been coined, despite the large number of problems which would be simplified by its use. (Cf. use of "conductance" in resistance problems.)

the technique described enables the time-scale to be expanded, theoretically up to infinity if required, without further increase in time-base voltage-gradient.

PART II: PRACTICAL LIMITATIONS

§ 1. INTRODUCTION

In the analysis given in Part I, several simplifying assumptions were made. Some of these will not be discussed in detail here: it will be assumed, for example, that the trace produced will be of sufficient intensity to be detected, and that synchronism can be maintained between the time-base generator and the phenomenon under investigation. The first of these requirements is somewhat eased if the phenomenon is repetitive in character (e.g. the R.F. wave-form of a pulsed magnetron). The second is perhaps rendered more stringent in the latter case, although it will often be possible to extract a suitable time-base voltage from the driving circuit.

Our present interest, however, is to consider the ways in which the deflection mechanism of the cathode-ray tube itself differs in practice from the ideal case. We shall also examine some possibilities which suggest themselves as offering at least partial solutions to the practical problems raised.

Four principal factors restrict the usefulness of the conventional cathode-ray tube deflection mechanism at high speeds. These are:

(a) "Intermodulation" between signal- and time-base plates, introducing distortion of the time-scale. (b) Mismatching of the signal circuit by the Y-deflection system, introducing wave-form distortion. (c) Transit-time effects, causing distortion of the frequency-characteristic of the Y-deflection system, and hence of any non-sinusoidal wave-form being examined. The linearity of the time-scale may also be affected. (d) Positional uncertainty caused by the finite cross-section of the beam at the deflection system. This produces an effect equivalent to defocusing. These will now be considered in turn.

§ 2. INTERMODULATION

Intermodulation or "cross-talk" between X- and Y-plates is caused chiefly by direct capacitive pick-up of the unwanted signal by the plates concerned. In its most troublesome form the higher-frequency components of the Y-signal find their way to the X-plates and may even cause loops in the trace. In this case, other things being equal, the effect is most serious at low time-base speeds, so that, for a given value of inter-X-plate impedance, the problem is eased by increasing the time-base voltage gradient. Clearly, since the stray capacitance and the inter-X-plate impedance form a potential divider, it is an advantage to employ a time-base generator with as low an output impedance as possible.

At very high frequencies, however, the inductance of the leads to the X-plates sets a lower limit to the inter-plate impedance, and it becomes essential to employ thorough screening to minimize X-Y-plate stray capacitance. Although the use of such screens is fairly common, the earthing arrangements require special care at high frequencies, as a shield earthed via a high-inductance lead forms an excellent coupling device.

Two forms of shield construction have been used in the course of work on a high-speed oscilloscope for the investigation of current build-up in radar modulator valves (Mackay 1946b). In both, a metal disc, slotted to permit

passage of the electron stream, was employed as a shield between X- and Y-plates. In one form of tube * the disc was made larger than the diameter of the tube neck, a copper-glass seal being employed, so that a continuous metallic shield could be arranged to enclose completely the portion of the tube containing the Y-plates and screen (figure 9).

The shielding obtained in this way was most satisfactory, but the production of a tube with a large glass-metal seal is not an easy problem. The second form of tube † employed was designed to obviate this. The slotted disc D (figure 10) had a diameter just less than that of the tube neck, and was fitted with a cylindrical collar C_1 of the same diameter. Outside the neck, an earthed "guard-ring" G coplanar with the disc formed a virtual extension of it, and was fitted with a second collar, corresponding in position to the first, and so acting as a capacitive by-pass of the earth lead from the disc. The latter was brought out to a thimble cap just behind the guard-ring.

In the same oscilloscope, the shielding provided by this tube also proved completely satisfactory up to the highest test frequency (600 Mc/s.) used.

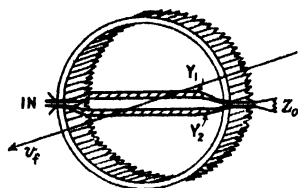


Figure 11.

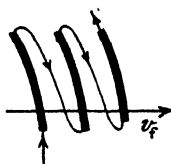


Figure 12.

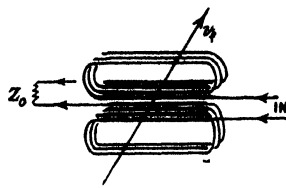


Figure 13.

Unfortunately work on the project was interrupted by the author's release from Admiralty service before tests at higher frequencies could be undertaken, but there seems no reason to doubt that this method of shielding could be satisfactory at frequencies several times higher than those used. The hard-valve time-base generator employed in these tests had an effective output capacitance of the order of 15 pF.

§ 3. MISMATCHING DISTORTION

In general the impedance presented by a deflector-plate system is capacitive, and since it is normally connected to the signal source via a cable with resistive or complex impedance, considerable distortion of the monitored wave-form can take place at high frequencies. Reflection of the signal can also introduce serious effects.

In an effort to overcome this difficulty, special Y-plate systems were incorporated in the tubes referred to in § 2. They were designed as sections of parallel-tape transmission line of the same impedance as the coupling cable, and leads were brought out from each side of the tube neck, so that the signal could travel *through* the deflector system to a second cable or a suitable resistive termination at the other side (figure 11). (A similar scheme has been developed independently in U.S.A. (Rudenberg 1945), with parallel wires as deflectors.) Apart from the small reflections occurring at discontinuities in impedance at the glass seals, such techniques can virtually eliminate loading effects.

* Produced in the research laboratories of Messrs. A. C. Cossor.

† Produced in E.M.I. Research Laboratories.

§ 4. TRANSIT-TIME

The effect of transit-time on X-deflection is discussed briefly in an Appendix. Its most serious result is the modification of the form of the functions V and U (Part I), particularly just as the deflecting voltage begins to rise. At this instant a number of electrons, already part-way through the field before the change took place, will emerge with disproportionately low values of v_x . This has the same effect as a rounding of the leading edge of the time-base wave-form. Since distortion of this type can be allowed for in the design of the Y-plates, as described in Part I, it does not represent a fundamental limiting factor.

In the case of the Y-plates, however, transit-time is of vital importance. A limit is placed on their effective narrowness not only by considerations of sensitivity and impedance (which are directly related quantities), but also by the edge effects which occur between narrow tapes. These effects can be reduced by a decrease in the plate separation, but unless micro-oscillographic technique (von Ardenne 1938, Hollmann 1940, Lee 1946) be adopted, the thickness of the electron beam normally restricts the usefulness of such a decrease. The use of a magnetic lens system, arranged to bring the beam to a subsidiary "cross-over" at the Y-plates, might however be of some assistance in reducing the permissible plate separation.

A suggestion for extending considerably the upper-frequency response of a deflector system has been put forward recently (Mackay 1946 a), and is illustrated in figure 12. Essentially the method proposed is to divide each Y-plate into a series of segments, each fed from the preceding one by way of a loop of conductor such that the signal takes the same time as an electron to travel from segment to segment. Thus each electron encounters the signal several times in succession, always in the correct phase, and receives cumulative deflections. In fact the arrangement is equivalent to a "travelling-wave" system with a velocity equal to that of the electron stream. Since v_e is normally of the order of $c/10$, it is necessary only to arrange that the signal path is about ten times the axial separation between successive segments. The "plates" might conveniently take the form of two distorted helices of D-shaped cross-section (figure 13), with earthed shields (not shown) so arranged as to maintain over the outer portions of the D's an impedance approximately equal to the characteristic impedance presented by the inner portions to the signal which is applied "in push-pull" to the input leads. The ends of the helices nearest to the screen are terminated in a suitable resistance, or may be connected to an output cable if required.

It is evident from Part I (§5) that an expansion of time-scale would be achieved if we could reduce the velocity of the signal along the path PQ (figure 7). It is possible in theory to do this by using a helically-wound conductor in place of each Y "plate" or wire, but this would tend to introduce slight irregularities in the sensitivity to Y-deflection, as a function of X-deflection.

§ 5. BEAM CROSS-SECTION

The cross-sectional dimensions of the stream of electrons (apart from their importance in defining the minimum plate separation) introduce a further limiting factor when the inclined-plate technique is used to increase the time-scale factor. This is best seen by reference to figure 14, in which AA'B' represents a magnified portion of the electron stream after X-deflection. The undeflected beam A'A

has a width a in the x, z plane, and since X-deflection is analogous to a shearing action, a typical cross-section LM of the kink A'B', normal to the tube axis, will also have a width a .

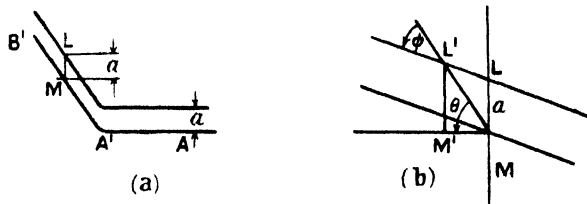


Figure 14.

Thus if conventional Y-plates of negligible axial length were employed, the width of the group of electrons LM receiving simultaneous deflections would be just a . If, however, the Y-plates are inclined at an angle θ to the tube axis, it is clear from figure 14(b) that the effective width L'M of the electron stream is increased to $a \cos(\theta - \phi) / \sin \phi$, where ϕ is the angle between the plates and the kink. Consequently in this case the picture on the screen will show a defocusing effect as if the spot were elongated in the X-direction, although it may appear perfectly circular when at rest. Assuming that the cross-sectional area of the beam is fixed by other considerations, it might be best in such a case to make a as small as possible, at the expense of the width in the Y-direction, so that the spot is normally elliptical. (This width a is that of the beam before it has been brought to a focus. The spot size is normally much smaller.)

Since in fact the Y-field will have a finite width in the axial direction, this must further increase the width of beam simultaneously receiving a given deflection. Indeed it may well prove that this problem is the chief obstacle to the realization of the full possibilities of any inclined-plate method of time-scale expansion. Factors of the order of $\times 10$ might, however, be usefully attainable with present-day tubes having beam-widths of the order of 2-3 mm. Further improvement may be possible with suitable tube design.

§ 6. BRIGHTNESS

We have hitherto ignored the question of trace brightness as outside the scope of this paper. It has been mentioned however (Part I, § 4) that the suggested method of scale expansion does not affect the brightness of the trace. The trace on the expanded time-scale is delineated by the same number of electrons as in the "conventional" case, so that the total illumination is independent of the degree of expansion obtained by inclining the Y-plates. The loss sustained is not in brightness, but in definition (§ 5).

Some indication of the possibilities with present-day materials was obtained in experiments conducted in 1946. At an accelerating voltage of 7 kv., and a repetition-rate of the order of 1000 c/s., the tubes referred to in § 2 gave a trace visible under a viewing hood at writing speeds up to and exceeding the velocity of light. (A hydrogen thyratron time-base generator circuit was employed.) This speed is adequate for the adoption of the inclined-plate technique (Part I, § 6).

§ 7. CONCLUSIONS

It has been shown that, in theory, any desired time-scale factor up to infinity can be attained in a C.R. oscillograph, by shaping the Y-plates to make use of the finite time taken by the beam to traverse the deflecting system. From examination

of the practical factors involved, the most useful lines to pursue towards this end would seem to be: (a) the reduction of beam cross-section at the Y-plates, to allow of a corresponding reduction in inter-plate separation, and to minimize "defocusing" effects; (b) design of the Y-plates as a parallel-wire transmission line of known characteristic impedance, preferably on a "travelling-wave" principle; (c) reduction of intermodulation effects to a minimum by means of efficient shielding. It is emphasized that although the theory outlined was developed in the course of oscillographic work at Admiralty Signal Establishment,* opportunities for practical investigations were limited to those few aspects which could be developed in the time available. Consequently the bulk of the present paper is presented only as a theoretical analysis of possibilities inherent in the oscillographic method, in the hope that further developments may be stimulated along the lines indicated.

ACKNOWLEDGMENTS

Part of the matter described forms the subject of Patent Applications made jointly with the Director of Scientific Research, Admiralty, and permission to publish the paper is gratefully acknowledged. The author also wishes to acknowledge the help and inspiration derived from discussions with many of his former colleagues in Admiralty Signal Establishment; he is particularly indebted to the able assistance of Mr. C. A. J. Clements in the experimental work.

APPENDIX

Effect of transit-time on X-deflection

The upper-limit effects of transit-time on X-deflection are most easily seen by considering the deflection produced by a "step" of voltage V_x , applied to X-plates of effective length l cm. (figure 1(a)). At the instant of application, a section of the beam of length l is between the plates, and it is the electrons in this section which will form the kink responsible for the trace on the screen.

Taking as temporary origin the point O where the electrons leave the field, electrons at distance ζ to the left of O will experience the field due to V_x for a time ζ/v_t .

Hence their transverse velocity v_x on leaving O is given by $v_x = kV_x\zeta/v_t$, and the kink (figure 7) therefore is parabolic with the equation

$$x = v_x(d - z)/v_t = kV_x z(d - z)/v_t^2,$$

since $z = \zeta$ when the origin is transferred to P.

Thus even an infinitely high time-base-voltage gradient is able to produce only a comparatively slowly-rising kink, whose slope depends mainly on the amplitude of the voltage step.

REFERENCES

- VON ARDENNE, M., 1938, *Z. Phys.*, **109**, 553.
 HOLLMANN, H. E., 1940, *Proc. Inst. Radio Engrs.*, **28**, 218.
 LEE, G. M., 1946, *Proc. Inst. Radio Engrs.*, **34**, 121.
 MACKAY, D. M., 1946 a, *Brit. Prov. Pat. Spec.* 18071/46; 1946 b, unpublished A.S.E. Report.
 RUDENBERG, H. G., 1945, *J. Appl. Phys.*, **16**, 279.

* Now Admiralty Signal and Radar Establishment.

Theoretical Design of Linear Accelerator for Electrons

By W. WALKINSHAW

Atomic Energy Research Establishment, Malvern

Communicated by J. D. Cockcroft; MS. received 17 March 1948

ABSTRACT. A circular corrugated waveguide is considered as an electron accelerator. The dependence of phase velocity on frequency and guide dimensions is obtained analytically and is in close agreement with experiment. Tables have been computed for designing accelerator tubes for any frequency and R.F. power.

§ 1. INTRODUCTION

A SHORT account of an experimental section of a travelling wave type of linear accelerator for electrons has been given in a recent paper (Fry, Harvie, Mullett and Walkinshaw 1947). It is our present purpose to describe in greater detail the theoretical analysis on which its design was based. Briefly, the accelerator tube consists of a circular waveguide propagating a wave with an axial component of electric field. Electrons, introduced in a monokinetic stream at the fed end of the guide, form bunches which "surf-ride" in an accelerating field and are maintained in phase with the wave by increasing the phase velocity at the same rate as the mean electron velocity. Ordinary laboratory waveguides cannot be used for this purpose, since their phase velocity is always greater than the velocity of light, but by making deep corrugations in the outer wall magnetic energy can be stored, and it is thus possible to decrease the phase velocity of the wave in a manner analogous to inductive loading of a transmission line in a delay network.

A first design requirement is a sufficiently accurate theory relating the phase velocity of the wave to the radio-frequency and the geometrical parameters of the corrugated guide. This is carried out and a most satisfactory agreement is found between theory and practice: for example, low power measurements made by Mullett and Loach (1948) at 3000 Mc/s. on a two-metre length of corrugated guide have shown a phase error of 18° at the end of the section; that is, an error of one part in four hundred in phase velocity. An account of this theory is given in § 2; numerical data are given in § 3 for the design of accelerator tubes.

Electrons are injected with an initial velocity considerably less than the velocity of light. The phase velocity of the wave is then increased slowly with distance at a rate depending on the available peak accelerating field. The mean acceleration rate must always be less than this peak accelerating field or the electrons fail to form stable bunches. To keep the accelerator length as short as possible some compromise must be made between the fraction of electrons trapped in the wave and the rapidity of the acceleration process. An analytical treatment of this part of the design will be found in § 3.

§ 2. UNIFORM CORRUGATED WAVEGUIDE

2.1. Frequency equation

Our purpose is to excite a travelling wave which has a phase velocity less than the velocity of light and an axial component of electric field. Let us examine therefore the simplest type of wave which has these properties, that is, a circularly

symmetrical solution of the electromagnetic field equations with axial and radial components of electric field and a circular magnetic component. Thus we get (Stratton 1941)

$$\left. \begin{aligned} E_\rho &= E_0(i\beta/\chi) J_1(\chi\rho) \exp[-i\beta z + i\omega t], \\ E_z &= E_0 J_0(\chi\rho) \exp[-i\beta z + i\omega t], \\ Z_0 H_\phi &= E_0(ik/\chi) J_1(\chi\rho) \exp[-i\beta z + i\omega t], \end{aligned} \right\} \dots\dots (1)$$

where ρ , ϕ and z are the usual cylindrical coordinates, Z_0 is the intrinsic impedance of free space, k is the wave number $= 2\pi/\text{free space wavelength}$, χ and β are the components of the phase constant in the ρ and z directions respectively and $k^2 = \chi^2 + \beta^2$.

When $\beta < k$, the phase velocity of this wave is greater than the velocity of light; χ is real, and from the properties of the Bessel function for real argument E_z has a maximum at $\rho = 0$ and varies in the usual sinusoidal fashion with increasing ρ . This is the type of solution familiar in ordinary laboratory waveguides in which the boundary surface coincides with a zero of E_z .

In the present application to electron acceleration the phase velocity is required less than the velocity of light; that is, $\beta > k$. χ is now imaginary and in accordance with the properties of the Bessel function for imaginary argument, E_z increases monotonically with increasing radial distance. It is apparent that a wave of this type cannot be contained within an ordinary smooth metallic surface. Examination of the radial field impedance, however, indicates the type of boundary surface which is required. The field impedance is

$$-E_z/H_\phi = Z_0(i\chi/k)\{J_0(\chi\rho)/J_1(\chi\rho)\}. \dots\dots (2)$$

When χ is imaginary this impedance is an inductive reactance for all values of ρ . Now, an inductive boundary surface can be simulated by making deep corrugations in a metallic surface to stop the axial energy flow in that region. Reactive energy is then stored inside the deep corrugations and by a suitable choice of depth any desired value of reactive impedance can be presented at the corrugation mouths.

The simplest type of wave which can exist inside each corrugation has an axial component of electric field and circular component of magnetic field. In other words the corrugation has the appearance of a short-circuited radial transmission line and will present any desired reactance at its mouth depending on its depth.

Let us now derive the relationship between guide dimensions, phase velocity and frequency. A crude approach, which is qualitatively accurate for guide wavelengths long compared with the corrugation width, assumes the field in the axial region given by equations (1) and equates the impedance derived in (2) to the impedance looking into the corrugation at its mouth. This approximation was first used by Cutler (1944).

A better approximation can be obtained, however, if we are able to make a sufficiently accurate estimation of the variation of tangential electric field at the mouths of the corrugations. It is found subsequently that the relationship between frequency and phase velocity is not critically dependent on this choice, hence a reasonable estimate of this field leads to a solution of high accuracy. We have therefore assumed the tangential field across the mouth of the n th corrugation at $\rho = a$ to be

$$E'_z = \{C \exp(-i\beta_0 nD)\} / \{1 - (2x_n/d)^2\}^{\frac{1}{2}}, \dots\dots (3)$$

and zero over the intervening metal walls. The symbols used have the following significance : a = radius of inner boundary of guide, b = radius of outer boundary of guide, D = corrugation pitch, i.e. distance between centre of corrugation walls, d = width of corrugation mouth, $\beta_0 = 2\pi/\lambda_g$, where λ_g is the guide wavelength. x_n is measured from the centre of the n th corrugation.

Now the general circularly symmetrical solution in the axial region of the guide must be represented by summation of solutions of the form derived in (1) by giving β (or χ) all possible values. But we see from (3) that the phase change per corrugation width is $\beta_0 D$, hence β can only take the values given by

$$\beta_m D = \beta_0 D + 2\pi m. \quad \dots\dots(4)$$

The complete solution for the field for $\rho < a$ is therefore

$$\left. \begin{aligned} E_\rho &= \sum_{-\infty}^{\infty} A_m (i\beta_m/\chi_m) J_1(\chi_m \rho) \exp[-i\beta_m z + i\omega t], \\ E_z &= \sum_{-\infty}^{\infty} A_m J_0(\chi_m \rho) \exp[-i\beta_m z + i\omega t], \\ Z_0 H_\phi &= \sum_{-\infty}^{\infty} A_m (ik/\chi_m) J_1(\chi_m \rho) \exp[-i\beta_m z + i\omega t], \end{aligned} \right\} \quad \dots\dots(5)$$

where β_m is given by (4) and $k^2 = \beta_m^2 + \chi_m^2$. The coefficients A_m are found in the usual manner of Fourier analysis such that, at $\rho = a$, E_z is given by (3). That is

$$\begin{aligned} A_m J_0(\chi_m a) D &= C \exp(-i\beta_0 n D) \int_{nD-d/2}^{nD+d/2} \frac{\exp(i\beta_m z)}{1 - (2x_n/d)^2} dz \\ &= (C\pi d/2) J_1(\beta_m d/2). \end{aligned} \quad \dots\dots(6)$$

Substituting this value for the coefficients in (5) gives the complete field in the region $\rho < a$, in accordance with the assumption made in (3).

We now turn our attention to the fields inside the corrugations. A similar analysis might be carried out to obtain a series solution for this field. Since, however, the corrugation width is small compared with the free space wavelength in the present application, only the transmission line mode is of any consequence; hence the fields in the n th corrugation are to a sufficient degree of accuracy given by

$$\left. \begin{aligned} E_z &= B F_0(k\rho) \exp(-i\beta_0 n D), \\ Z_0 H_\phi &= i B F_1(k\rho) \exp(-i\beta_0 n D), \end{aligned} \right\} \quad \dots\dots(7)$$

where $B = C\pi/2F_0(ka)$,

$$\left. \begin{aligned} F_0(ka) &= J_0(ka) Y_0(kb) - Y_0(ka) J_0(kb), \\ F_1(ka) &= J_1(ka) Y_0(kb) - Y_1(ka) J_0(kb), \end{aligned} \right\} \quad \dots\dots(8)$$

and b is the radius of the outer surface of the guide.

The required frequency equation is now obtained from the condition that the magnetic fields given by (5) and (7) must match at their common boundary. Since this can be done accurately at only one point in the corrugation mouth we equate the magnetic fields averaged over the opening. This results in the frequency equation

$$\frac{F_1(ka)}{F_0(ka)} = \frac{d}{D} \sum_{-\infty}^{\infty} (ik/\chi_m) \frac{J_1(\chi_m a)}{J_0(\chi_m a)} J_0(\beta_m d/2) \left\{ \frac{\sin(\beta_m d/2)}{\beta_m d/2} \right\}. \quad \dots\dots(9)$$

This is the required relationship between guide dimensions, wave velocity and frequency and we shall discuss its accuracy later. Let us return now to a qualitative description of the physical implications of the foregoing analysis.

The fields inside the corrugations require no further elucidation. To our approximation these are simple radial transmission line solutions with a constant phase shift in the field between adjacent sections. Since an inductive impedance is required at the corrugation mouth the depth must be slightly less than an effective quarter-wave transmission line.

In the axial region of the guide, the field is given qualitatively by the first term ($m=0$) of the series given by (5), when the phase change of the field per corrugation is large. For phase velocities less than the velocity of light, the field in the inner region is concentrated close to the corrugation mouths, falling off towards the guide centre in exponential fashion in the nature of a surface wave. For the acceleration of slow particles this is an undesirable feature of the corrugated guide. Of particular interest for electron acceleration is the case when the phase velocity equals the velocity of light. Then $\beta = k$ and making $\chi \rightarrow 0$ reduces (1) to

$$\left. \begin{aligned} E_\rho &= \frac{1}{2} i E_0 k \rho \exp(-ikz + i\omega t) \\ E_z &= E_0 \exp(-ikz + i\omega t) \\ Z_0 H_\phi &= \frac{1}{2} E_0 k \rho \exp(-ikz + i\omega t) \end{aligned} \right\} \dots\dots (10)$$

The axial electric field is then independent of its distance from the guide centre. It is also of interest to note that the impedance of this wave is equal to the intrinsic impedance of free space, that is, the wave impedance of a plane wave in free space.

As in unloaded waveguides, there is a low frequency cut-off which is approximately equal to the cut-off for an ordinary guide having a radius equal to the outer radius of the corrugated guide. In addition there is also a high frequency cut-off which is peculiar to transmission systems having a periodic structure. In general the higher components of the field given by (5) are only important close to the corrugation mouths. However, when the frequency is increased the phase velocity decreases until ultimately the phase change per corrugation width is 180° . From (4) we see that $\beta_{-m-1} = -\beta_m$ or, in other words, a standing wave is produced in the guide. For a travelling wave system it will be necessary to use a sufficient number of corrugations per wavelength to avoid this resonance condition.

2.2. *Experimental check of frequency equation*

The theoretical analysis described above was carried out in conjunction with an extensive investigation of corrugated guides by Mullett and Loach (1948). For a detailed comparison of theory and experiment reference should be made to their paper. Suffice it to say here that approximations of varying degrees of accuracy were examined and discarded in favour of the above theory, which was considered sufficiently accurate (Harvie 1948) for design purposes.

2.3. *Axial field at guide centre*

So far our analysis has given a satisfactory solution for phase velocity. Equal in importance, in the design of an accelerating tube, is an accurate estimate of the axial field at the guide centre. Assuming for the moment that only the

fundamental component ($m=0$) is of importance, the Poynting vector for the power flow can be integrated in the normal way across the inner cross-section of the guide. We get for the axial field at the guide centre

$$E_0 = \left\{ \frac{2WZ_0\chi^2}{\pi a^2 k \beta \{J_1^2(\chi a) - J_0(\chi a)J_2(\chi a)\}} \right\}^{\frac{1}{2}} \dots\dots(11)$$

This is shown in figure 1 as a function of phase velocity for various values of inner radius. It will be noticed that as the phase velocity is decreased the field at the centre also decreases because the major portion of the power flow then takes place in the region near the mouths of the corrugations. For the same reason the

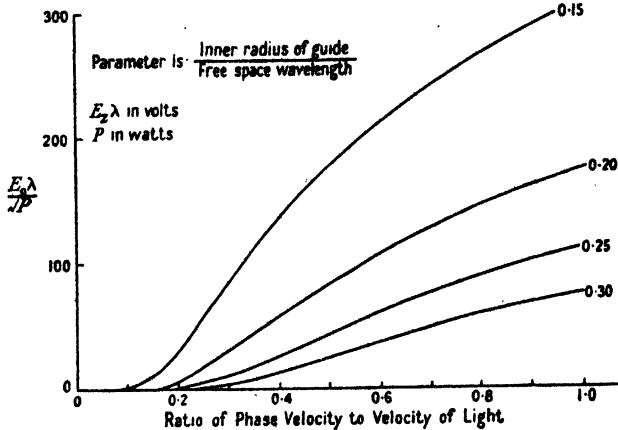


Figure 1. Dependence of axial electric field on phase velocity and inner guide radius.

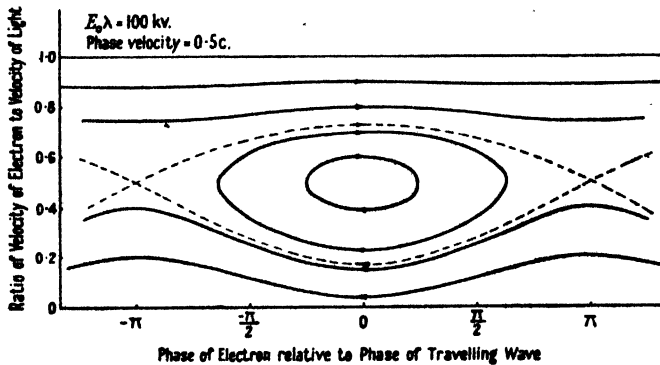


Figure 2. Electron motion in uniform travelling w.a.e.

power flow due to the higher components of the field is only important close to the corrugations. A rough calculation has shown that for a guide of inner radius 0.2λ , with ten corrugations per free space wavelength and a phase velocity 0.4 times the velocity of light, approximately 98% of the power is carried by the lowest component ($m=0$).

The field at the centre of the guide is therefore given to a sufficient degree of accuracy by (9) for guides likely to be used for linear accelerators.

When the phase velocity is equal to the velocity of light, (11) reduces to

$$E_0 = (480W)^{\frac{1}{2}}(\lambda/\pi a^2) \dots\dots(12)$$

Equations (9) and (11) are the basic equations, which we shall use in later sections for design purposes.

§ 3. DESIGN OF LINEAR ACCELERATOR TUBE

3.1. Motion of electrons in travelling wave

We now turn our attention to the motion of electrons injected into a travelling wave; in particular, we consider the early stages of an accelerator when the electron velocity is appreciably different from the velocity of light. We recall that it is our intention to introduce electrons into the travelling wave and increase the phase velocity in such a way that stable bunches of electrons are formed slightly in advance of the peak accelerating field. In general the equations of motion for an electron inside a circular corrugated guide are insoluble, but in order to arrive at suitable design data for the variation of phase velocity along the accelerator, we shall assume that electrons are constrained to the axis. In fact electrons bunch at a position of radial defocusing and a paraxial magnetic field is required to confine the electrons to the axial region. The coil design for this field has been carried out by Harvie (1948) and will not be discussed here.

It is instructive to consider first the simple case of an electron in a travelling wave with constant phase velocity. The equation of motion for the constrained electron of charge e is

$$d(m\dot{z})/dt = -Ee \sin(\beta z - \omega t) \quad \dots\dots(13)$$

where m = relativistic mass of particle, $\beta = 2\pi/\text{guide wavelength}$, ω/β = phase velocity of wave, E = amplitude of electric field. Multiplying either side of (13) by $(\dot{z} - \omega/\beta)$ and integrating with respect to time t gives the energy equation

$$mc^2[1 - \omega\dot{z}/\beta c^2] = (Ee/\beta) \cos(\beta z - \omega t) + \text{constant}. \quad \dots\dots(14)$$

This can be written in the equivalent form

$$E_1/E_0 - (\omega/c\beta)[(E_1/E_0)^2 - 1]^{\frac{1}{2}} = (E/\beta E_0)[\cos \beta y + \text{constant}] \quad \dots\dots(15)$$

where E_1 = sum of rest and kinetic energies, E_0 = rest energy, $y = z - \omega t/\beta$; βy is the phase of the particle relative to a position of constant phase in the travelling wave where the field is zero.

The cosine term on the right-hand side indicates that the variation of energy with phase is periodic. This is demonstrated by a typical case in figure 2. Those electrons injected with a velocity approximately equal to the wave velocity subsequently oscillate around a stable phase position of zero field. Fast electrons progress continuously forward whilst slow electrons fall continuously backwards. In neither of these cases does an electron gain any energy per cycle. In order to do this it is necessary to start the wave off with a velocity almost equal to that of the injected electrons and slowly increase its phase velocity. The stable phase position now falls back towards the peak of the accelerating field, its position depending on the mean rate at which the electrons are accelerated. Let us suppose then that the phase velocity of the wave is increased in such a way that an electron injected at a phase $-A$, say, is maintained at this phase throughout its subsequent motion. That is, if the amplitude of the field is constant, this electron is accelerated by a constant field $E \sin A$. Its energy equation is therefore

$$mc^2 = Eex \sin A + \text{constant}. \quad \dots\dots(16)$$

This relationship gives the required phase velocity of the wave, which is of course equal to the velocity of this electron.

It is of further interest to examine the subsequent motion of those electrons which enter the guide at phases other than $-A$. For this purpose we write the differential equation of motion of the electron as

$$d(m\dot{z})/dt = -Ee \sin [f(z) - \omega t], \quad \dots\dots(17)$$

where $f(z)$ is the phase change of the wave along the guide, and is already determined (except for a constant) by the constraint put on the stable electron. That is

$$f(z) - \omega t = -A. \quad \dots\dots(18)$$

Integration of (16) gives a relationship between z and t which when substituted in (18) gives

$$f(z) = \frac{\omega m_0 c}{Ee \sin A} \left[\left\{ \frac{Ee \sin A}{m_0 c^2} z + 1 \right\}^2 - 1 \right]^{\frac{1}{2}} + \text{constant}. \quad \dots\dots(19)$$

By numerical integrations of (17) with $f(z)$ given by (19), the phase oscillations of electrons introduced with a velocity of $0.4c$ into a guide commencing with this wave velocity, are shown in figure 3 (a) for various initial phase angles. The

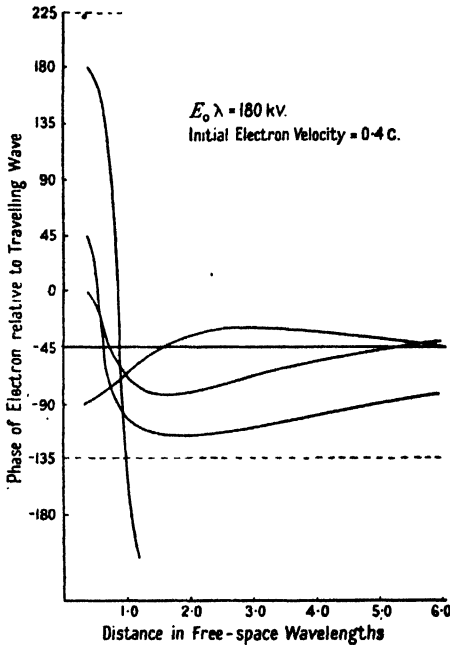


Figure 3 (a). Variation of electron phase with distance.

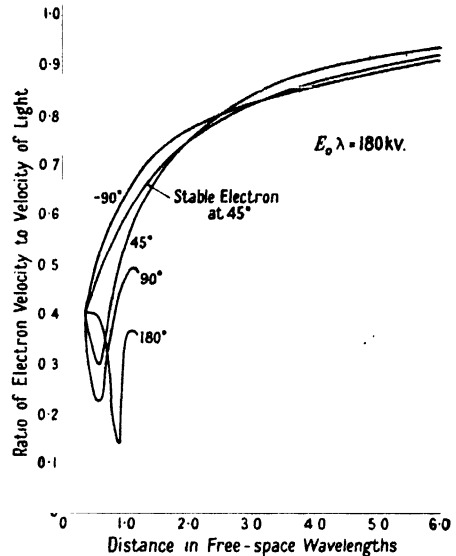


Figure 3 (b). Variation of electron velocity with distance.

amplitude of the wave is 0.18 mv. per free space wavelength and the stable phase position is 45° in advance of the peak accelerating field. The corresponding variation in velocity of these electrons is shown in figure 3 (b). Those electrons entering near the stable phase position oscillate with decreasing amplitude as their energy increases. If, however, an electron is introduced too far in advance of the stable phase, as for example the electron initially at 180° , it fails to receive sufficient energy for it to be trapped by the wave. Only those electrons initially between -135° and roughly 60° are raised to higher energy levels with a mean velocity equal to that of the wave. It is apparent that, if the stable phase position had been close to the peak accelerating field, the mean rate of acceleration would be greater, but fewer electrons would be trapped.

3.2. Length of tapered section

For energies above one or two megavolts, equation (15), for a travelling wave with constant phase velocity, can be written as

$$E_1/E_0(1 - \omega/c\beta) + \frac{1}{2}(\omega E_0/c\beta E_1) = (E/\beta E_0)(\cos \beta y + c). \quad \dots (20)$$

If now the wave velocity, ω/β , is taken equal to the velocity of light, c , and it is required that electrons of initial energy E_1 remain within $\pm 10^\circ$ of the peak accelerating field in attaining extremely high energies, then the following inequality must hold :

$$E_1 > \pi E_0^2 / (2E\lambda \cos 80^\circ) = (2.3/E\lambda). \quad \dots (21)$$

For high energy accelerators $E\lambda$ is likely to be at least 1 MV., hence tapering of the tube can cease for electron energies above 3 MV. without fear of serious phase slip.

3.3. Numerical design data

Equations (9), (11) and (16) are the basic equations which we shall now use to obtain numerical data for designing accelerator tubes. It will be noted that the amplitude of the wave at the guide centre has been assumed independent of distance in the previous section. We see from equation (11) or figure 1 that this can only be achieved for different phase velocities by varying the inner radius of the guide. In the following table guide dimensions are given as a function of phase velocity such that the field at the centre is 175 kv. per free space wavelength for a transmitted power of 1 MW. For a phase velocity equal to that of light the inner radius is the 0.2 times the free space wavelength.

Table 1. Phase velocity versus guide dimensions

$D/\lambda=0.1$; $d/D=0.75$; $E\lambda=175\sqrt{W}$ volts at $\rho=0$; W in watts

Phase velocity	Inner radius		Outer radius	
Velocity of light	Free space wavelength		Free space wavelength	
		δ_2^*		δ_2^*
0.4	0.136025	-3086	0.41814	-180
0.5	0.150615	-2137	0.42166	-110
0.6	0.162992	-1549	0.42403	-67
0.7	0.173779	-1166	0.42569	-45
0.8	0.183376	-905	0.42688	-34
0.9	0.192053	-720	0.42772	-27
1.0	0.200000	-585	0.42830	-23

* δ_2 in the above table is the modified second difference.

The above table can be used to design an accelerator tube for any arbitrary power and any chosen variation of phase velocity with distance. In particular the 40-centimetre tapered guide described in a separate paper (Fry, Harvie, Mullett and Walkinshaw 1947) was constructed on the assumption that a power of one megawatt was available, and such that the stable phase was 45° in advance of the peak accelerating field. The dimensions of this guide are easily obtained from (16) in conjunction with table 1, and we shall not give these in detail. The predicted behaviour of the electrons in this guide has already been shown in figures 3(a) and 3(b).

We shall conclude this section with two other tables similar to table 1. The first of these is for a guide differing from that just discussed in corrugation pitch and wall thickness and subsequently in corrugation depth. The inner radius is the same, that is, 0.2 times the free space wavelength for a phase velocity equal

to the velocity of light and such that constant peak field is maintained at the guide centre for other phase velocities. Owing to the high-frequency cut-off it is inadvisable to reduce the number of corrugations per guide wavelength to less than four or five, hence only phase velocities near the velocity of light are of interest.

Table 2. Phase velocity versus guide dimensions

$D/\lambda=0.2$; $d/D=0.875$; $R\lambda=175\sqrt{W}$ volts at $\rho=0$; W in watts

Phase velocity	Inner radius		Outer radius
Velocity of light	Free space wavelength		Free space wavelength
		δ_a	
0.8	0.183376	—905	0.4248
0.9	0.192053	—720	0.4244
1.0	0.200000	—585	0.4237
1.1	0.207356		0.4229

The final table is for a high peak field accelerator. Harvie (1948) has shown that the efficiency of the linear accelerator increases with decreasing inner diameter. Engineering tolerances and frequency sensitivity of the guide set a limit to its smallness, but experience has shown that a smaller radius than 0.2 times the free space wavelength for a phase velocity equal to that of light can be used. The following table is for a guide which has a peak accelerating field of 500 kv. per free space wavelength per megawatt transmitted power. The efficient length of a single section of this accelerator tube at a frequency of 3000 Mc/s. is likely to be of the order of two metres.

Table 3. Phase velocity versus guide dimensions

$D/\lambda=0.1$; $d/D=0.75$; $E\lambda=500\sqrt{W}$ volts at $\rho=0$; W in watts

Phase velocity	Inner radius		Outer radius
Velocity of light	Free space wavelength		Free space wavelength
		δ_a	
0.4	0.08791	—175	0.3956
0.5	0.09501	—114	0.3964
0.6	0.10090	—80	0.3969
0.7	0.10597	—59	0.3974
0.8	0.11044	—45	0.3978
0.9	0.11445	—36	0.3982
1.0	0.11810	—29	0.3985

ACKNOWLEDGMENTS

The work on linear accelerators has been the result of the combined efforts of a large number of people and the author wishes to acknowledge collectively the experimental work and ideas of his colleagues in addition to those specifically referred to already. This work was carried out by the Electronics Group of the Atomic Energy Research Establishment, Malvern, and the author wishes to thank the Director for permission to publish these results and acknowledge his indebtedness to Miss R. Hensman and Miss M. Kingston for their assistance in the computations.

REFERENCES

- CUTLER, C. C., 1944, *Electromagnetic Waves Guided by Corrugated Conducting Surfaces*, CRB 45/593 Bell Telephone Lab.
 FRY, D. W., HARVIE, R. B. R-S-, MULLETT, L. B., and WALKINSHAW, W., 1947, *Nature, Lond.*, **160**, 351.
 HARVIE, R. B. R-S-, 1948, *Proc. Phys. Soc.*, **61**, 255.
 MULLETT, L. B., and LOACH, B. G., 1948, *Proc. Phys. Soc.*, **61**, 271.
 STRATTON, J. A., 1941, *Electromagnetic Theory* (New York: McGraw-Hill), p. 360.

Travelling Wave Linear Accelerators

By R. B. R-SHERSBY-HARVIE

Atomic Energy Research Establishment, Harwell

Communicated by J. D. Cockcroft; MS. received 15 April 1948

ABSTRACT. Acceleration of electrons by means of an electromagnetic wave travelling down a circular corrugated tube is considered. The relevant properties of corrugated tubes are derived and the factors influencing electron energies. Some numerical results are also quoted for a wavelength of 10 cm. The problems of high energy accelerators are discussed briefly and illustrated by calculations which have been made for a 500 mev. accelerator.

§ 1. INTRODUCTION

PRELIMINARY investigations were started on the type of accelerator which is about to be described towards the end of 1945. Successful linear accelerators using radio-frequency accelerating fields had been previously reported (Sloan and co-workers 1931, 1934), and it was apparent that as a result of wartime advances in centimetre-wave magnetrons and waveguide technique, linear accelerators compared favourably with other forms of particle accelerators. Work began on electron accelerators, electrons being chosen both because of the technical simplicity associated with them and because there were important practical applications for comparatively low energies, 20 to 50 mev., in the field of medical therapy. The type first investigated was that in which electrons are accelerated by electromagnetic waves travelling down a circular corrugated waveguide; the successful operation of an experimental model has been reported by Fry and co-workers (1947).

The object of the present paper is to set out the main factors influencing the design of a travelling wave accelerator. Although it is not proposed to make detailed comparisons with other types of linear accelerator, mention will be made of some other methods which are available. The most important and obvious alternative is to accelerate the electrons by passing them through a series of resonant cavities excited from a radio-frequency source. Successful experiments using this method have been reported from several sources (Bowen, Pulley and Gooden 1946, Halpern *et al.* 1946, Allen and Symonds 1947) and other more ambitious projects are being worked upon (Schiff 1946, Alvarez 1946, Haxby *et al.* 1946, Akeley 1946, Morrison 1947, Schultz *et al.* 1947). It appears that these methods can yield energies comparable with the travelling-wave system, given the same amount of radio-frequency energy and the same overall length of accelerator (Walkinshaw 1947). There are, however, important practical differences not fully resolved, and neither type has as yet been proved superior to the other.

Another type of combined travelling-wave and resonator accelerator using a rectangular corrugated guide has been investigated by Mullett (1946) but was not pursued, mainly on account of difficulties of manufacture of the necessary equipment. Travelling-wave systems using various configurations of bent waveguides have also been considered and rejected for reasons of difficulty with focusing, practical inconvenience or sometimes inefficient accelerating performance, for example the coiled-up waveguide (Hudspeth 1946). Other variations of these methods and a proposal to use a circular waveguide lined with a dielectric tube have also been put forward (Willshaw 1945).

The circular corrugated waveguide travelling-wave type accelerator has at least not been found inferior to other methods for the acceleration of electrons; it has been shown (Cowhig 1946) that different methods must be employed for the acceleration of particles at velocities much less than that of light.

§ 2. DESCRIPTION.

The basic principles of the travelling-wave accelerator will be described briefly to save reference to other papers.

An electromagnetic wave with a longitudinal electric field component is set up in a corrugated tube; this is a circularly symmetrical wave with a field configuration such as is shown in section in figure 1. If a bunch of electrons situated at X is travelling at a velocity equal to the phase velocity of the wave, the electrons will be subjected to an electric field tending to accelerate them. This acceleration can be continued indefinitely, provided the phase velocity of the wave follows the velocity of the electrons as they accelerate; this is ensured by suitable dimensions of the corrugated tube.

If the bunch of electrons be situated at a phase angle θ with respect to the wave, the position of the bunches will be longitudinally stable provided $0 < \theta < \pi/2$, but will tend to diffuse away from the axis of the tube. Thus some form of focusing must be used; this has so far taken the form of a steady longitudinal magnetic field, the effect of which is discussed in Appendix C. Electrons are injected along the axis in a continuous stream and become bunched, due to the natural longitudinal stability of the accelerating process (Walkinshaw 1948).

Details of the practical construction of an accelerator are given by Fry and co-workers (1947), Mullett (1947), and Mullett and Loach (1948).

§ 3. SOME PROPERTIES OF THE CORRUGATED WAVEGUIDE

3.1. General

The performance of an accelerator is mainly determined by the properties of the corrugated waveguide. For a general view of the subject it is only necessary to consider these properties for the case where the phase velocity is constant and equal to c , the velocity of light; the velocity of an electron is

$$u = c\{1 - (m_0 c^2/E)^2\}^{1/2}, \quad \dots\dots(1)$$

where m_0 = rest mass of electron, E = total energy of electron; hence, since the rest energy $m_0 c^2$ is approximately 0.51 mev., the velocity of the electrons will be very close to that of light except for a short length near the start of an accelerator.

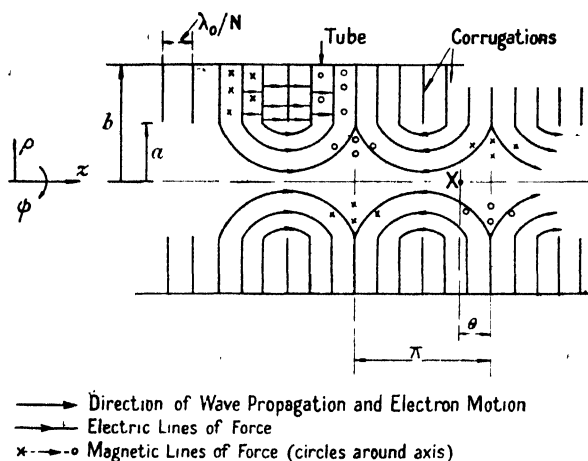


Figure 1. Diagrammatic view of fields in corrugated tube for the case where $v = c$.

This initial part, where the velocity is changing rapidly, is vital to the satisfactory working of the accelerator—the waveguide design for this part is dealt with by Walkinshaw (1948)—but the energies obtainable and the general characteristics are almost entirely determined by the main length over which the electrons and the wave will have a velocity very near to that of light. Consequently it will be assumed henceforth that the phase velocity of the wave is actually equal to the velocity of light.

3.2. The fields in the waveguide

The fields in the corrugated waveguide can be very simply calculated by a method due to Cutler (1944). This method gives results of sufficient accuracy for the present argument, although more refined calculations (Walkinshaw 1948) are necessary to design the waveguide to the high degree of precision required in an actual accelerator. Cutler assumes that the protuberances on the guide walls are very thin, and very close together compared with the wavelength. Hence the walls may be considered to have a uniform impedance (Booker 1947), giving simple boundary conditions. The fields in the guide are derived in Appendix A. The following are some special features of the wave when $v = c$: (i) the accelerating field, E_z , is uniform over the open part of the tube, so that electrons which are not properly aligned on the axis will still receive the same accelerating field; (ii) equation (A 7) shows that, with the same amount of power flowing in each case, the accelerating field is inversely proportional to the square of the inner radius; (iii) as the inner radius decreases, the fields inside the corrugations tend to those in an ordinary TM_{010} resonator, and, for the small inner radii used, the fields are not very different from those in the outer portions of such resonators.

3.3. Attenuation of the waveguide

It will be assumed that the electronic current passing down the guide is so small that power absorption due to it is negligible, and that the sole cause of attenuation is absorption of the wave by the walls of the tube. This may not be true in certain practical cases, but the effect is readily allowed for in cases of large electron currents, and the small current case is of sufficient importance to justify its exclusive study.

To calculate absorption by the walls it will be assumed that the fields remain the same as those already given, even when a finite corrugation spacing is used. Walkinshaw (1947) has shown that it is not desirable to use corrugation spacings exceeding about one-fifth of a wavelength, i.e. $N \approx 5$ in figure 1, for which this assumption is justified. The usual methods (Schelkunoff 1943) are used for finding the attenuation in waveguides.

The attenuation constant is

$$\alpha = w/2W \text{ nepers/metre,} \quad (2)$$

where W = power flow past any point in the guide, and w = power per metre absorbed by walls at the same point. The power absorbed is given by

$$w = (R/2) \iint H.H^* ds, \quad (3)$$

where H = magnetic field on the metal surfaces of the guide, H^* = complex conjugate of H , R = intrinsic resistance of the metal = $\sqrt{(\pi\mu c/\lambda_0 g)}$, μ = permeability of metal, g = conductivity of metal. The integration is carried out over the whole

surface of the metal boundaries over a unit length of pipe, the fields being assumed the same as in the absence of attenuation. It follows from (2) and (3) that the attenuation can be written

$$\alpha = \alpha_0 \lambda_0^{-3/2}, \quad \dots\dots(4)$$

where α_0 is a function of the properties of the metal and the geometry of the guide. When the phase velocity is fixed (e.g. $v = c$), the geometry is specified by a/λ_0 , the ratio of the inner radius of the guide to the wavelength, and N , the number of corrugations per wavelength.

Most of the loss is found to occur in the plane dividing walls between the corrugations, so that the attenuation is roughly proportional to N ; the variation with a/λ_0 is not so simple and is best found numerically. The attenuation has been computed for a copper tube with $N=10$ and with $v = c$, and the result is shown in figure 2, which gives α_0 as a function of a/λ_0 .

It has been found by experiment (Mullett 1948) that the measured attenuations of the guides which have been made have been roughly twice the calculated values. Thus the theoretical attenuation for $N=10$ may be taken as approximately equal to the practical attenuation for $N=5$, which approaches the optimum design of guide.

3.4 Electron energy obtainable

The energy gained by an electron in travelling down the tube may now be calculated. The maximum energy gained in a length z is

$$V(\text{ev.}) = \int_0^z |E_z| dz. \quad \dots\dots(5)$$

This is a maximum because it assumes that the electron travels all the time in the maximum accelerating field. This is not quite attainable in practice because insufficient electrons would be captured at the start of the accelerator (Walkinshaw 1948) and because it is not a position of stable phase equilibrium. However, experiments with the 0.5 mev. accelerator (Mullett 1947) have shown that about half the possible electron output can be obtained at a phase corresponding to an accelerating field greater than 90% of the maximum; under these conditions the output current may be restored to its full value by injecting more electrons at the beginning. This, together with the possibility of varying the phase along the accelerator, indicates that (5) over-estimates the attainable energies by not more than 10%. No correction will be included in the results, but the effect should not be forgotten.

The power travelling down the tube is

$$W = W_0 \exp(-2\alpha z), \quad \dots\dots(6)$$

where W_0 = power fed into the beginning of the waveguide at $z = 0$. Substituting

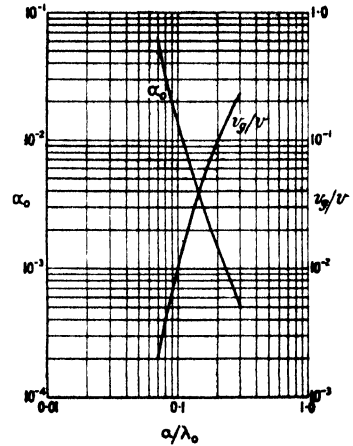


Figure 2. Attenuation and group velocity in corrugated tube.

Attenuation constant is $\alpha = \alpha_0 \lambda_0^{-3/2}$ nepers/metre.

Curve gives α_0 for the case where $N=10$ corrugations/wavelength, $v = c = 3 \times 10^8$ metres/sec., $g = 5.8 \times 10^7$ mhos/metre (typical copper).

Group velocity (v_g) is for the case where $v = c$.

from (6) and (A8) into (5) and integrating, $V \simeq 6.98(\lambda_0/a^2)\sqrt{W_0}\{1 - \exp(-\alpha z)\}/\alpha$, whence from (4)

$$V/\sqrt{(\lambda_0 W_0)} \simeq 6.98(\lambda_0/a^2)\{1 - \exp(-\alpha_0 z \lambda_0^{-3/2})\}/\alpha_0. \quad \dots\dots(7)$$

This relationship has been computed numerically, and figure 3 shows $V/\sqrt{(\lambda_0 W_0)}$ as a function of $z\lambda_0^{-3/2}$ for four different values of (a/λ_0) and for $v=c$ and $N=10$. To give an indication of the magnitudes involved, the same diagram has been plotted for the particular case where $\lambda_0=10$ cm. and $W_0=2 \times 10^6$ watts, and is shown in figure 4. The envelope of the family of curves, a dotted line in figures 3 and 4, shows the maximum attainable energy for any given length of accelerating tube, each length corresponding to a definite value of (a/λ_0) .

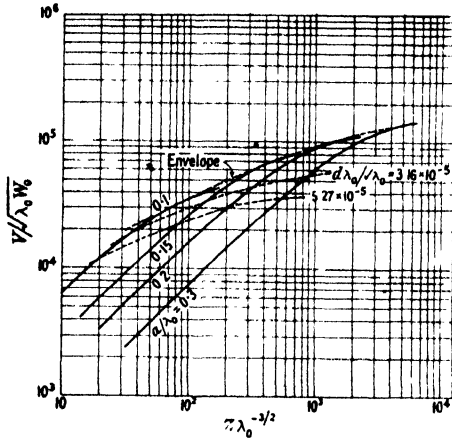


Figure 3. Electron energy.

Computed energy for the case where
 $N=10$ corrugations/wavelength,
 $v=c=3 \times 10^8$ metres/sec.,
 $g=5.8 \times 10^7$ mhos/metre (copper),
 λ_0 =free space wavelength (metres),
 z =distance along tube (metres),
 V =electron energy (ev.),
 a =inside radius of tube (metres),
 W_0 =radio wavelength power (watts).

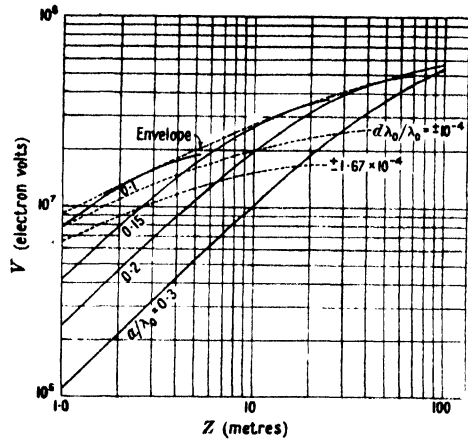


Figure 4.

Energy obtainable for the case where
 $\lambda_0=0.1$ metre,
 $W_0=2 \times 10^6$ watts,
 $N=10$ corrugations/wavelength,
 $v=c=3 \times 10^8$ metres/sec.,
 $g=5.8 \times 10^7$ mhos/metre.

It should be noted that the energy limitation represented by the envelope curve is entirely due to attenuation; other, often more serious, limitations will be discussed later. It can be shown* that, using the optimum value of a/λ_0 , the attainable electron energy is about 90% of that in the ideal, but hypothetical, case where all the power is dissipated in the tube and the accelerating field is uniform along the length. The conditions on the envelope curve are very close to those which give this 90% of the ideal energy. Thus it will be seen that the system of acceleration is not inherently inefficient in spite of the progressive falling off in accelerating field along the tube and the power remaining at the end, which has to be absorbed in an artificial load. Improvement could be effected by such means as reducing a/λ_0 as the end of the accelerator is approached, but the advantage is too small for such complications to be practicable.

* The problem is formally identical to that of determining the effect of aperture distribution in linear aerial arrays, and has been studied in detail in this connection.

3.5. Effect of wavelength and temperature variation, and group velocity

In an ideal accelerator the phase velocity of the wave would remain unchanged as the applied frequency is altered. In a corrugated waveguide this is by no means the case, and the phase velocity changes rapidly with applied frequency because a large amount of energy is stored in the corrugations and the guide acts as a dispersive medium. This is expressed by means of the concept of group velocity (Appendix B), and figure 2 shows v_g/v , plotted as a function of a/λ_0 .

If the waveguide is designed to have the correct phase velocity at a certain wavelength, then altering this wavelength will change the phase velocity, and the particles being accelerated will fail to remain in the correct phase relationship with the wave. Due to the low group velocity in the guide, this effect can be very serious. To calculate the phasing errors which occur it will be convenient to assume that the particle velocity remains unchanged, which will be very approximately true for particles having a high energy and a velocity approaching that of light.

The phase of a particle in the wave, at position z and time t , is $\phi = \omega t - 2\pi z/\lambda_g$. If the particle velocity is constant, and equal to u , this becomes

$$\phi = 2\pi(z/\lambda_0)\{c/u - \lambda_0/\lambda_g\}.$$

Hence

$$d\phi = 2\pi \frac{z}{\lambda_0} \frac{d\lambda_0}{\lambda_0} \left\{ \frac{\lambda_0}{\lambda_g} \frac{v}{v_g} - \frac{c}{u} \right\}.$$

For the case of interest, where $v = u = c$, this can be written

$$\frac{d\lambda_0}{\lambda_0} \sqrt{\lambda_0} = \frac{d\phi}{2\pi z \lambda_0^{-3/2} (v/v_g - 1)}. \quad \dots\dots (8)$$

Thus if a certain permissible phase error, $d\phi$, is chosen, contours of the corresponding permissible wavelength error, $\sqrt{\lambda_0} d\lambda_0/\lambda_0$, can be drawn on figures 3 and 4. Two such contours are shown for the case of a phase error of $2\pi/10$ radians. The arbitrarily chosen error of $2\pi/10$ is probably rather larger than can be tolerated in practice, and so the results shown are a conservative estimate of the extremely high wavelength stability which is required. The inherent instability of most practical sources would impose very serious limitations on the performance, and it will usually be necessary to incorporate some device for adjusting the wavelength of the source. By this means it should be possible to counteract all but the most rapid variations of wavelength.

Thermal expansion of the waveguide has practically the same effect as the same proportional change of wavelength, but in the opposite sense. Consider a uniform linear expansion of the waveguide accompanied by the same proportionate increase of wavelength. The wavelength in the guide, λ_g , will increase in the same proportion and the phase velocity will remain unchanged; hence the phasing of the electrons with respect to the wave will be unaffected. If the wavelength is now restored to its original value, the effect of the change will be the same as if only the linear expansion had taken place. Hence, for small changes, the phase errors caused by a uniform linear expansion are the same as those caused by the same proportionate change of wavelength in the opposite sense.

The coefficient of expansion of copper is about 1.6×10^{-5} . Thus a change of 1°C. in guide temperature is equivalent to a wavelength change $d\lambda_0/\lambda_0 = -1.6 \times 10^{-5}$. In figure 4, therefore, the wavelength tolerance contours of $d\lambda_0/\lambda_0 = \pm 10^{-4}$ and $\pm 1.67 \times 10^{-4}$ are equivalent to temperature tolerances of $\mp 6^\circ \text{C.}$ and $\mp 10^\circ \text{C.}$ respectively.

The actual phase errors in any accelerator will be the sum of the errors caused by wavelength and temperature changes. Where only wavelength tolerances are given these should be considered as representing the sum of both tolerances.

The practical effect of temperature changes is not nearly so serious as that of wavelength changes because they occur slowly, and any uniform change can be compensated by altering the wavelength of the power source. Thus it is only necessary to reduce temperature differences between different parts of the waveguide, and the effect of any remaining temperature differences will certainly be less than that of the same uniform temperature changes, which have already been calculated.

The foregoing estimate of the permissible wavelength variations has been based upon the assumption of a single frequency of oscillation. In practice, linear accelerators must usually be operated from pulsed sources in order to obtain sufficient radio-frequency power; this involves the application of a whole spectrum of frequencies. The steady-state stabilities give a first approximation to those required under pulsed conditions, but various anomalies and difficulties arise in connection with the use of very short pulses, and the problem requires further investigation. The minimum pulse duration which can be used is governed by the time required to build up the wave in the corrugated tube; to a first approximation this is equal to the length of tube divided by the group velocity. The accelerator will only be operative during that portion of the applied pulse which occurs after the wave has had time to build up.

3.6. *Dimensional tolerances*

The effect of one class of dimensional variation—thermal expansion—has already been considered and, from the results which were obtained, it will have been evident that the waveguide dimensions must be so accurately determined that no normal workshop process could be used to manufacture the guide. It also appears most unlikely, from the results of experiment (Mullett and Loach 1948), that the guide could be designed with sufficient accuracy, even were the former limitation removed, except perhaps by long and costly experiment. However, these difficulties can be circumvented to some extent, and it is necessary to discuss the point in more detail, because it has an important bearing upon the design of an accelerator.

It has already been shown that the phase velocity of the wave must correspond to the electron velocity so closely that the relative phase difference between electron and wave does not become excessive. Nevertheless relatively large errors in phase velocity are permissible locally provided they are not cumulative. Thus an accelerator could be constructed from sections of guide with relatively large velocity errors if some subsequent means of phase adjustment can be provided between the sections. It is anticipated that this can be done by interposing between sections short lengths of corrugated guide, which can be made or selected from a stock. This procedure should be easily applicable, provided the number of adjustment points which are required does not become excessive. It is also advisable to distinguish between systematic and accidental errors in the guide; the former are probably mainly due to design and the latter to construction.

Little can be said of the systematic errors, since it is not possible to predict their variation with the shape of the guide (a/λ_0). Among the experimental

results of Mullett and Loach (1948) on a guide with $a/\lambda_0 = 0.2$ and $\lambda_0 = 10$ cm. is one showing phase errors of about 18° in a 2-metre length which appear to be mainly systematic errors. It is only possible to assume that in guides with different values of a/λ_0 the proportional dimensional errors will remain the same; it should be possible to reach this standard with a reasonable amount of experiment. On this basis it follows that the number of adjustment points will remain constant in accelerators with constant wavelength tolerance $d\lambda_0/\lambda_0$ and wavelength λ_0 , but with varying a/λ_0 and total length. It also follows that the spacing between adjustment points is proportional to λ_0 if a/λ_0 remains constant. The experimental result quoted shows a phase error of the order of that tolerable, and it can thus be said that adjustment points should be spaced about 2 metres apart in a guide with $a/\lambda_0 = 0.2$ and $\lambda_0 = 10$ cm. On this assumption a guide with $a/\lambda_0 = 0.1$ and $\lambda_0 = 10$ cm. would require adjustment points at about 20 cm. spacings.

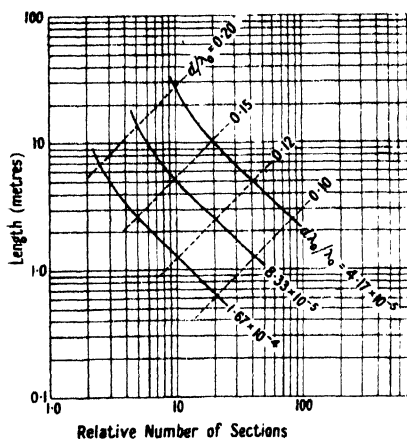


Figure 5.

Relative number of sections required for adjustment of random errors for the case where $\lambda_0 = 0.10$ metres. $v = c = 3 \times 10^8$ metres/sec.

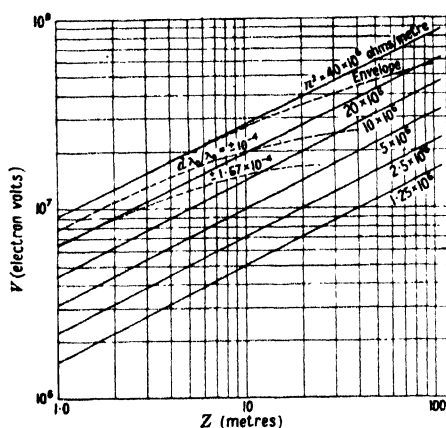


Figure 6.

Efficiency of accelerator of figure 4.
Parameter $\pi^2 = \text{efficiency}$.

The accidental errors are assumed to be caused by random errors in the dimensions of each cavity in the corrugated guide. These will cause a certain phase error, and the phase errors in each cavity are supposed to be capable of addition to give the total error in a length of guide. Since the sum of such a number of independent errors gives a resultant error whose mean is proportional to the square root of the number of additions made, the mean phase error in a section of waveguide will be proportional to the square root of its length, in contrast to the case of systematic errors where the resultant phase error was taken as proportional to the length. Over a moderate range of sizes, workshop practice is such that the proportional dimensional errors may be assumed constant. Thus the length of sections between the adjustment points required to compensate for random errors will be proportional to the wavelength λ_0 and can be computed as a function of a/λ_0 by means of the group velocity as already discussed in connection with thermal expansion. This type of computation has been carried out,

and figure 5 shows the relative number of adjustment sections required as a function of length of accelerator for different values of wavelength tolerance and for the case $\lambda_0 = 10$ cm. and $N = 10$. The absolute number of sections required is not determinate at present through lack of experimental evidence, upon which estimates of errors must of course be based. However, although it is known that the accidental errors are in practice negligible compared with the systematic for $a/\lambda_0 = 0.2$, it is clear that the effect of the former increases much more rapidly as a/λ_0 is reduced. Estimates of probable tolerances indicate that the random errors are almost certain to predominate for the case $a/\lambda_0 = 0.1$, $\lambda_0 = 10$ cm. and $N = 10$.

It will be shown in § 5 that there are certain advantages to be gained by reducing a/λ_0 as far as possible. One limit to this process is the resulting increase in waveguide errors and the number of adjustments required to compensate for them. The tolerable number of adjustments is a practical limitation, but it could be very serious if, for example, it were necessary to adjust separately each corrugation cavity. It is hoped that in due course further experimental results will become available to enable absolute values to be given to the errors.

Two favourable factors have not been mentioned. Firstly, increasing the corrugation pitch (i.e. reducing N) may reduce the accidental errors by reducing the number of cavities, and it has already been mentioned that pitches corresponding to $N \simeq 5$ are desirable. Secondly, a certain amount of compensation for cumulative errors may sometimes be possible by adjusting the wavelength of the power source.

§ 4. EFFICIENCY OF ACCELERATORS

Accelerator performance is affected by the alteration of various waveguide design parameters; help in choice of such parameters is obtained by consideration of concept of accelerator "efficiency", a figure of merit, defined as π^2 , where $\pi^2 = V^2/W_0 L$ ohms/metre; V and W_0 have the same meanings as before and L = length of accelerator in metres. The efficiency figure has several useful properties. Since both power and length are expensive in cost and space, high efficiency is an indication of economy. Again, high energy accelerators will usually have to be constructed by operating a number of smaller accelerators in cascade. It can be shown that the efficiency of such an accelerator is the same as that of its component accelerators, so that a knowledge of the efficiency of the components enables the overall length and power requirements to be rapidly assessed; for the same reasons the efficiency is an excellent basis of comparison even between accelerators of widely different design, although it does not take into account wavelength differences and many other features of practical importance. Physically, the efficiency figure is almost the same as the "dynamic resistance" of a unit length of the accelerator, and it is often useful to think of it as such.

Constant efficiency contours may be plotted on diagrams such as figures 3 and 4 (see figure 6). The constant efficiency curves are simply straight lines with the efficiency proportional to the square of the intercept on the ordinate. It will be seen at once that the efficiency increases as a/λ_0 is reduced, although a limiting maximum value is eventually reached, and that this increase of efficiency is more marked in accelerators with a wide wavelength tolerance. Hence there is good reason for attempting to use accelerators made up of a number of sections

with small a/λ_0 rather than of a smaller number of sections with large a/λ_0 ; either the total length or total power will be reduced thereby.

The efficiencies obtainable with the corrugated guide appear very low compared with that of the Sloan and Coates accelerator (Sloan 1934), which was some 15-20 times higher; this is primarily due to the fact that the Sloan and Coates accelerator was designed to accelerate heavy ions, moving very much more slowly than light, so that the field in a cavity resonator (or other resonant circuit) may be used for acceleration several times without introducing much additional loss.

§ 5. GENERAL DESIGN PROCEDURE

There are so many variables to take into account that it is impossible to lay down any rigid rules of design; it is best to set down the most important considerations and leave details to be determined for any specific case.

One of the first things to decide upon in designing an accelerator is the radio wavelength to be used. As the wavelength is reduced, the efficiency increases slowly (the precise law depends upon the wavelength tolerance and a/λ_0 chosen) and the mass of metal in the waveguide and vacuum system decreases. On the other hand, more power is available from radio valves at longer wavelengths and, while magnetrons can be paralleled (Cowhig and Jones 1947), this may be an important point if a very high energy accelerator is required. Against this, however, is the fact that large tubes store more energy, so that longer pulses are required at longer wavelengths: roughly speaking, the build-up time is proportional to wavelength. Less important points are that wavelength reduction enables the size of focusing coils and their power consumption to be reduced, although too short a wavelength may not allow a sufficient aperture for the passage of electrons. In general it appears that as short a wavelength as possible is desirable, consistent with obtaining the necessary power.

For a given wavelength and output energy, the approximate efficiency obtainable is known in terms of the product of length and power. The balance between length and power is mainly one of economy and individual requirements. Short accelerators using high power also require stronger focusing fields and more focusing power than longer ones using less radio-frequency power.

The last important choice is probably that of the value of a/λ_0 . Efficiency demands small a/λ_0 , although there is not much to be gained by going below a certain value. Besides the additional difficulty involved in construction of the guide when a/λ_0 is reduced, and the resultant increase in the number of phase adjustment points required, the number of separate component accelerators comprised in the whole accelerator is increased and the power has to be split up and fed in at a greater number of points. It is necessary to arrive at a compromise for a long accelerator composed of several sections, but, if the object is to produce the shortest possible accelerator, it will be necessary to reduce a/λ_0 as far as possible. The difficulties increase very rapidly as a/λ_0 is reduced, so that there is a sharp limit to the smallest guide which can be made, although this limit is not yet known; tubes with a/λ_0 as small as 0.15, at a wavelength of 10 cm., have been made without difficulty at Telecommunications Research Establishment. It is estimated that the limit probably lies in the region of $a/\lambda_0 = 0.1$.

None of the factors mentioned above is of course unaffected by the others, although there is a certain degree of independence. As an example of design

procedure some data are reproduced which were computed by Taylor (1946) to investigate the construction of a 500 mev. accelerator. The length has been fixed arbitrarily and figure 7 shows the power required, as a function of the number of component accelerator sections, for three different wavelengths and wavelength tolerances. Many other facts require consideration, but this chart shows clearly the effect of some of the variables and the magnitudes involved. It should perhaps be added that no decision has been made to construct such an accelerator.

§ 6. ADDITIONAL REMARKS

The electron energy obtainable from accelerators has been discussed comprehensively, but little has been said about the beam current obtainable. There are two important limitations to the current-space charge and damping of the wave. Nothing useful is known about the former, but attention is drawn to the fact that the experimental 0.5 mev. accelerator passed currents of up to about 100 ma. during the pulse. The latter limitation merely means that no more electronic energy can be got out of the device than the radio-frequency put in. Thus it should not be difficult to design linear accelerators giving relatively large mean or peak currents, provided sufficient power is available. Indeed, there seems no reason why high energy efficiencies should not be attainable.

There are three major advantages where exceedingly high energies are involved (greater than, say, 1000 mev): the fact that the cost is roughly proportional to the electronic output energy required, which appears to be so far a unique feature of linear accelerators, the small amount of power lost by radiation, and the short electron path, as compared with orbital machines, which reduces the vacuum problems. Also, in all linear accelerators the actual electrons are readily led out of the machine.

The construction of two further accelerators is at present contemplated. The first, which is nearly completed, is effectively an extension of the former 0.5 mev. machine. It is designed to give 4 mev. and has the specification: length = 2 metres; power = 2 MW.; wavelength = 10 cm.; $a/\lambda_0 = 0.2$; injection energy = 50 kev. Its main object is to investigate focusing problems in longer accelerators than the 0.5 mev. machine. The second is a 20 mev. accelerator of the same length, but much higher efficiency (small a/λ_0).

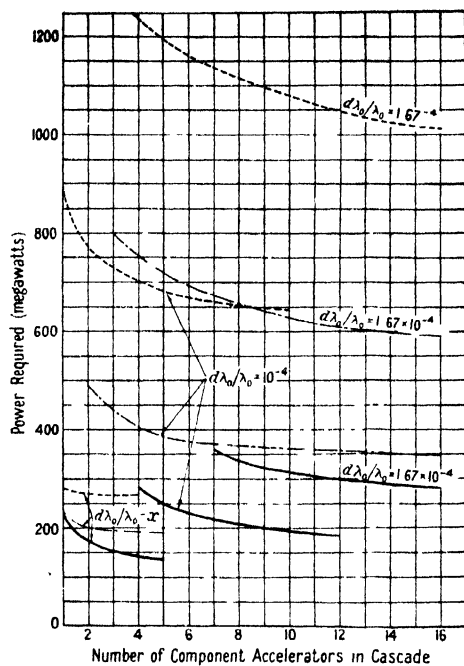


Figure 7. Accelerator for 500 M v

Total length = 50 metres.

— — — λ_0 0.10 metres.

- - - - λ_0 0.25 metres.

- - λ_0 0.50 metres

Frequency tolerance x is such that energy is limited by attenuation.

ACKNOWLEDGMENTS

The work on linear accelerators has been the result of the combined efforts of a number of people, and the writer wishes to acknowledge collectively the experimental results and ideas due to his colleagues. This work was carried out as part of the programme of the Atomic Energy Research Establishment, and acknowledgment is made to the Director for permission to publish these results.

APPENDIX A

FIELDS IN THE CORRUGATED TUBE

Consider a waveguide with dimensions and cylindrical coordinate system as shown in figure 1. Using rationalized M.K.S. units, the wave which is of interest can be written, for the inner part of the tube, as

$$E_z = -iE_0J_0(\chi\rho); \quad E_\rho = E_0(\beta/\chi)J_1(\chi\rho); \quad Z_0H_\phi = E_0(k/\chi)J_1(\chi\rho), \quad \dots\dots (A1)$$

where λ_0 = wavelength in free space (m.), λ_g = wavelength in corrugated tube (m.), $k = 2\pi/\lambda_0$, $\beta = 2\pi/\lambda_g$, $\chi = k[1 - c/v^2]^{1/2}$, v = phase velocity of wave in m/sec., Z_0 = intrinsic impedance of free space in ohms, $= (\mu_0/\epsilon_0)^{1/2} \simeq 120\pi$, J_n = first kind of Bessel function of order n , E_0 = arbitrary constant.

The factor $\exp i(\omega t - \beta z)$ is implied for each field component; the losses are assumed negligible for the moment.

It will be seen that for phase velocities greater than light this wave is the same as the familiar TM_{01} wave used in smooth-walled tubes (Schelkunoff 1943).

Again the fields within the corrugations can be written (Schelkunoff 1943)

$$E'_z = BJ_0(k\rho) + CY_0(k\rho); \quad Z_0H'_\phi = iBJ_1(k\rho) + iCY_1(k\rho). \quad \dots\dots (A2)$$

At the boundary where $\rho = b$ it is necessary to have $E'_z = 0$. Hence from (A2) it follows that

$$BJ_0(kb) + CY_0(kb) = 0. \quad \dots\dots (A3)$$

Thus the radial impedance looking into the corrugations, $Z_\rho = -E'_z/H'_\phi$, is given by

$$Z_\rho = iZ_0 \frac{Y_0(kb)J_0(ka) - J_0(kb)Y_0(ka)}{Y_0(kb)J_1(ka) - J_0(kb)Y_1(ka)}. \quad \dots\dots (A4)$$

To satisfy the boundary conditions at the surface of the corrugations the radial impedance of the wave in the inner part of the tube must be the same as that in the corrugations. Thus from (A1) this impedance can also be written

$$Z_0 = iZ_0(\chi/k)J_0(\chi a)/J_1(\chi a). \quad \dots\dots (A5)$$

Simultaneous solutions of (A4) and (A5) give waveguide dimensions in terms of phase velocity or vice versa. Examination shows that any phase velocity can be obtained by a suitable choice of dimensions.

The strength of the fields in the corrugations can be related to those in the centre part by noting that $E'_z = E_z$ and $H'_\phi = H_\phi$ when $\rho = a$.

It is also necessary to know the value of the constants in terms of the power flowing down the tube. Integrating the Poynting vector over the inner portion of the tube,

$$2W = E_0^2(k/\chi)^2(c/v)(\pi a^2/Z_0)\{J_1^2(\chi a) - J_0(\chi a)J_2(\chi a)\}, \quad \dots\dots (A6)$$

where W = power flowing down tube in watts.

The special case where $v = c$ is of particular importance for present purposes. Substituting this value of v in (A1) and (A6), it is found that the longitudinal field is given by

$$E_z = -i(4/ka^2)\sqrt{(WZ_0/\pi)} \text{ volts/metre,} \quad \dots\dots (A7)$$

or, substituting the values of constants, $E_z \simeq -i(6.98/a^2)\lambda_0\sqrt{W}$.

APPENDIX B

GROUP VELOCITY IN THE TUBE

The group velocity in a waveguide can be defined as (Lamont 1942)

$$v_g = c \left(\frac{\lambda_g}{\lambda_0} \right)^2 \frac{d\lambda_0}{d\lambda_g} = \frac{v^2 d\lambda_0}{c d\lambda_g}. \quad \dots\dots (B1)$$

A method of finding the group velocity is described which it is believed gives a clearer physical picture of the relevant factors than a direct derivation from $d\lambda_0/d\lambda_g$. The results also emerge in a form well adapted for computation from the integrals required in computing the attenuation, and show how the attenuation and group velocity are related.

Results will be used which have been obtained in connection with the deformation of electromagnetic cavity resonators, as given, for example, by MacLean (1945).

Consider a portion of a loss-free corrugated waveguide formed into a resonator with a loosely fitting piston at one end as shown in figure 8. The corrugations must also be loosely fitting, so that they can be pushed along by the piston, and they are again assumed indefinitely thin and closely spaced. Let the cavity be set oscillating in the desired mode

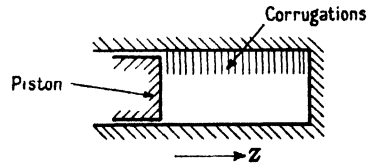


Figure 8.

and let the length be equal to one wavelength, λ_g , of this mode. Let the piston now be slowly pushed in a small distance, dz , while the cavity continues to perform free oscillations. Then if F be the time-averaged force on the piston due to the electromagnetic fields, the work done on the piston is $F dz = \text{increase of oscillating energy} = d\mathcal{E}$, say. But it has been shown (MacLean 1945) that $\mathcal{E}\lambda_0 = \text{constant}$, or $d\lambda_0/\lambda_0 = -d\mathcal{E}/\mathcal{E}$. Hence, since $dz = -d\lambda_g$, substitution in (B1) gives

$$v_g = vF/(\mathcal{E}/\lambda_g). \quad \dots\dots (B2)$$

The expression $\mathcal{E}/\lambda_g$ is of course just the electromagnetic energy per unit length.

To evaluate (B2) two well-known relationships in electromagnetic theory are used: the time-averaged force on a perfectly conducting surface, reflecting an electromagnetic wave, $F = \frac{\mu}{4} \iint H \cdot H^* ds - \frac{\epsilon}{4} \iint E \cdot E^* ds$ newtons, and the time-averaged energy stored in a volume, $\mathcal{E} = \frac{\mu}{4} \iiint H \cdot H^* dv + \frac{\epsilon}{4} \iiint E \cdot E^* dv$ joules. Substituting these in (B2) and using the field equations (A1), there results

$$v/v_g = (v/c)^2 \left\{ 1 + \frac{\iint H'_\varphi \cdot H'^*_\varphi ds_2}{\iint H_\varphi \cdot H^*_\varphi ds_1} \right\}, \quad \dots\dots (B3)$$

where the integral in the numerator is across the corrugated section of the guide and that in the denominator is across the open part. It will be seen that the

integral in the numerator is proportional to the losses in the dividing walls of the corrugations—which is the major part of the total losses—while that in the denominator is proportional to the power flow down the guide. Thus v/v_g is very nearly proportional to the attenuation. v_g/v has been computed for the case $v = c$, and the result is shown in figure 2.

It is interesting to notice that an observer moving down the guide at the group velocity v_g will observe no net energy flow past him, the forward flow of energy in the centre part being balanced by the backward flow in the corrugations which are moving past the hypothetical observer. Details of the appearance of waveguide fields to moving observers have been published (Hershberger 1945). This confirms that the group velocity is in this case also the velocity of energy flow, and indicates another method of deriving the group velocity.

APPENDIX C

FOCUSING

The need for focusing

(i) Consider a positively charged particle travelling down the waveguide parallel to the axis with velocity v . Let v be also the phase velocity of the electromagnetic wave at the same point as the particle. Gaussian (mixed) units will be used.

The particle is subjected to an outward radial force given by $F_p = eE_p - H_q e v/c$, where e is the charge on the particle. But from waveguide theory $H_q = (v/c)E_p$; therefore

$$F_p = eE_p\{1 - (v/c)^2\}. \quad \dots\dots(C1)$$

From waveguide theory it again follows that E_p is positive in the bunching and accelerating region of the wave (i.e. the region where E_z is positive and $\partial E_z/\partial z$ is negative). Hence, from (C1), F_p is positive in this region and the guide is diffusing. In fact E_p is always of the opposite sign to $\partial E_z/\partial z$.

(ii) Consider a similar case where the particle velocity is u but the wave velocity is still v . Then, arguing as before, we find that

$$F_p = eE_p\{1 - uv/c^2\}. \quad \dots\dots(C2)$$

Hence in this case the guide would only be focusing in the bunching region if $uv > c^2$; the wave would then have a greater velocity than the particle, which would involve phase changing at intervals, but the method provides a form of focusing not obtainable with a realizable form of magnetic field, and might be valuable in a long high-energy accelerator. It is not practicable to make use of this form of focusing unless the particle velocity is fairly close to c .

Magnetic focusing

It is possible to prevent the particles being diffused by immersing the accelerator in a steady longitudinal magnetic field. Ideally this field must be so designed that the particles are guided down the waveguide with the minimum expenditure of energy in the focusing magnet. This problem has not been solved to date, and we content ourselves with examining the field required to produce a particular form of guiding.

We assume that particles are to be guided down the accelerator in a helical path of small radius, r , centred on the axis of the accelerator, and consider the

case of an "equilibrium" particle, i.e. one in the stable phase position, and with constant phase angle.

Let H_z and H_ρ be the components of a circularly symmetrical steady magnetic focusing field and let the particle position be (ρ, ϕ, z) .

Then the equations of motion are $d(m\rho^2\dot{\phi})/dt = -(e/c)\rho\dot{H}_z + (e/c)\rho\dot{z}H_\rho$, which on integration gives

$$m r^2 \dot{\phi} \simeq -\frac{e}{c} \frac{r^2}{2} H_z + C, \quad \dots\dots (C3)$$

where C is a constant, which we shall discuss later, and

$$d(m\dot{\rho})/dt - m\rho\dot{\phi}^2 = eE_\rho - (e/c)\dot{z}H_\phi + (e/c)\rho\dot{\phi}H_z.$$

Now since $H_\phi = (\dot{z}/c)E_\rho$ and $\dot{\rho} = 0$, we find from these equations

$$H_z^2 = 4 \left[\left(\frac{c}{e} \right)^2 \left(\frac{C}{r^2} \right)^2 + \frac{m c^2}{e} \frac{E_\rho}{r} \left\{ 1 - \left(\frac{\dot{z}}{c} \right)^2 \right\} \right].$$

Now, except at the beginning of an accelerator, it is nearly true to say that the particle energy is the same as if its velocity were $\dot{z}$. Hence we get

$$H_z^2 = 4 \left[\left(\frac{c}{e} \right)^2 \left(\frac{C}{r^2} \right)^2 + \left(\frac{m_0 c^2}{e} \right)^2 \frac{e E_\rho}{r} \frac{1}{m c^2} \right]. \quad \dots\dots (C4)$$

Thus (C4) gives the longitudinal component of the required field. It will be seen that it is desirable that C should be finite in order to secure stability, but not too large, as this would lead to unnecessary expenditure of focusing energy in the later stages of an accelerator; space charge should also contribute towards stability, but its effect is not easily estimated. For purposes of practical computation C is assumed negligible.

Note that if C is finite the focusing field H_z never vanishes. There are two explanations for this, depending upon the sign of C . In one case the particles eventually follow a path parallel to the accelerator axis without any angular momentum but in a field of finite H_z ; although there is no focusing action, removal of H_z would itself create angular momentum. In the other case there is a residual angular momentum, which requires the focusing field to be continued.

In a practical accelerator the above formula for the focusing field does not hold near the beginning of the accelerator, since the electrons are not easily injected in bunches or in the proper direction to satisfy the conditions postulated. Calculations of the focusing field have been made for other special cases, corresponding to different electron trajectories, and the results have been surprisingly similar to the above. It is therefore concluded that the electron injection arrangements are not critical and that it is unlikely that any other trajectory can be found which would require substantially less focusing field.

Extension to practical cases

It has so far been the custom to design accelerating waveguides so that the peak longitudinal field on the axis, E_0 , remains constant. It is therefore better to write (C4) in terms of E_0 and other convenient quantities.

From waveguide theory we have $\hat{E}_\rho = E_0(\beta/\chi)J_1(\chi\rho)$, where $\beta = 2\pi/\lambda_g$, $\chi = (2\pi/\lambda_0)\{1 - (c/v)^2\}^{1/2}$, and $v = \dot{z}$ = phase velocity of wave.

Assuming as before that the particle velocity is approximately $\dot{z}$, and putting $E_p = \hat{E}_p \cos \theta$, where θ is the particle-to-wave phase angle, we have from (C4)

$$H_z^2 = 4 \left[\left(\frac{c}{e} \right)^2 \left(\frac{C}{r^2} \right)^2 + \frac{m_0 c^2}{e} \frac{E_0 \cos \theta}{r} I_1 \left(\frac{2\pi r}{\lambda_0} \frac{c}{\dot{z}} \frac{m_0}{m} \right) \right].$$

We can put $C \simeq 0$; $I_1 \left(\frac{2\pi r}{\lambda_0} \frac{c}{\dot{z}} \frac{m_0}{m} \right) \simeq \frac{1}{2} \left(\frac{2\pi r}{\lambda_0} \frac{c}{\dot{z}} \frac{m_0}{m} \right),$

whence, writing the energy in terms of electron volts, we obtain as a practical formula

$$H_z^2 = 4(m_0 c^2/e)(E_0 \cos \theta)(\pi c/\lambda_0 \dot{z})V_0/(V + V_0). \quad \dots\dots (C5)$$

Example

Formula (C5) has been used to compute the longitudinal axial component of the magnetic focusing induction required for the 0.5 mev. accelerator already referred to. Figure 9 shows B_z as a function of z for a number of different values of the radio-frequency input power W_0 . It should be noted that changing W_0 changes both the strength of the radio-frequency fields and the stable phase angle θ , so that the focusing field changes more rapidly than it would if due to only one of these effects. For $W_0 = 0.5$ MW. the focusing field is zero.

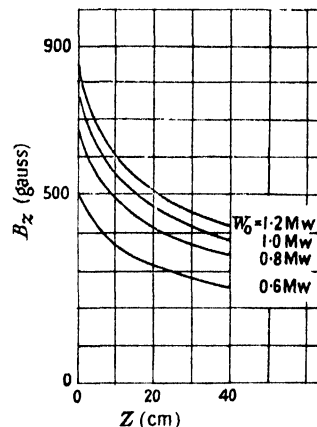


Figure 9.

REFERENCES

- AKELEY, E. S., 1946, *Phys. Rev.*, **69**, 50, 255, 1946, *J. Appl. Phys.*, **17**, 1056.
 ALLEN, W. D., and SYMONDS, J. L., 1947, *Proc. Phys. Soc.*, **59**, 622.
 ALVAREZ, L. W., 1946, *Phys. Rev.*, **70**, 799.
 BOOKER, H. G., 1947, *J. Instn. Elect. Engrs.*, **94**, pt. iii, 171.
 BOWEN, E. G., PULLEY, O. O., and GOODEN, J. S., 1946, *Nature, Lond.*, **157**, 840.
 COWHIG, W. T., 1946, *T.R.E. Memo.*, S1/M178/WTC.
 COWHIG, W. T., and JONES, E. J., 1947, *T.R.E. Report*, A.E. 1008.
 CUTLER, C. C., 1944, *Bell Tel. Labs. Report*, MM-44-160-218.
 FRY, D. W., R.-SHERSBY-HARVIE, R. B., MULLETT, L. B., and WALKINSHAW, W., 1947, *Nature, Lond.*, **160**, 351.
 HALPERN, J., EVERHART, E., RAPUANO, R. A., and SLATER, J. C., 1946, *Phys. Rev.*, **69**, 688.
 HAXBY, R. O., et al., 1946, *Phys. Rev.* **70**, 797.
 HERSHBERGER, W. D., 1945, *J. Appl. Phys.*, **16**, 465.
 HUDSPETH, E. L., 1946, *Phys. Rev.*, **69**, 671.
 LAMONT, H. R. L., 1942, *Wave Guides* (London: Methuen & Co.).
 MACLEAN, W. R., 1945, *Proc. Inst. Radio Engrs.*, **33**, 539.
 MORRISON, P., 1947, *J. Appl. Phys.*, **18**, 133.
 MULLETT, L. B., 1946, *T.R.E. Report*, T2021; 1947, *Ibid.*, AE1006, AE1007.
 MULLETT, L. B., and LOACH, B. G., 1948, *Proc. Phys. Soc.* **61**, 271.
 SCHELKUNOFF, S. A., 1943, *Electromagnetic Waves* (New York: D. Van Nostrand).
 SCHIFF, L. I., 1946, *Rev. Sci. Instrum.* **17**, 6.
 SCHULTZ, H. L., et al., 1947, *Phys. Rev.*, **72**, 346.
 SLOAN, D. H., and COATES, W. M., 1934, *Phys. Rev.*, **47**, 539.
 SLOAN, D. H., and LAWRENCE, E. O., 1931, *Phys. Rev.*, **38**, 2021.
 TAYLOR, H. E., 1946, *T.R.E. Report*, A.E.1000.
 WALKINSHAW, W., 1947, *T.R.E. Report*, A.E. 1005; 1948, *Proc. Phys. Soc.*, **61**, 246.
 WILLSHAW, W. E., and LAMONT, H. R. L., 1945, *Confidential Report from Research Labs. of General Electric Co. Ltd.*, No. 8766(B).

Experimental Work on Corrugated Waveguides and Associated Components for Linear Electron Accelerators

BY L. B. MULLETT AND B. G. LOACH

Atomic Energy Research Establishment, Harwell

Communicated by J. D. Cockcroft; MS. received 17 March 1948

ABSTRACT. This paper describes experimental work carried out on corrugated waveguides in connection with the development of an electron linear accelerator.

The general problem of feeding R.F. power from a conventional rectangular waveguide is considered and several successful "feeds" are described.

Finally, the work carried out on corrugated circular waveguides with continuously varying phase velocities and their associated "feeds" for the 0.5 mev. and 4 mev. sections of long accelerators is described in some detail. It is shown that the experimental results are in close agreement with the theoretical treatment of Harvie, Cutler and Walkinshaw.

§ 1. INTRODUCTION

THE development of the linear accelerator in the form described by Harvie (1948), involved a considerable programme of experimental work on corrugated (or loaded) waveguides. From the wide variety of possible modes and geometrical systems, the corrugated circular waveguide carrying an E_{01} type of mode was very soon chosen as the most likely basis for a travelling-wave type of accelerator. One other possible geometry was considered—a rectangular waveguide loaded on one of the broad faces with rectangular corrugations somewhat longer than a half-guide wavelength—and a certain amount of theoretical and experimental work was carried out on this system, although this work was later abandoned in favour of the circular corrugated waveguide system.

The programme of experimental work was initially directed towards testing the accuracy of various theoretical treatments (of ever increasing complexity) in order to establish a means of designing systems with predetermined laws relating phase velocity and accelerating field to distance.

The design wavelength is 10.00 cm. throughout.

§ 2. CONSTRUCTION

A mechanical form taken by a corrugated circular waveguide is shown in figures 1 and 2. Separate sections are machined from pure copper with a general tolerance of ± 0.0005 in. on all internal dimensions. These sections are self located by spigots and are soft-soldered together on the outside as shown in figure 2. Other forms of construction are being considered. Each section has four $3/16$ in. diameter holes spaced 90° apart both for measurement purposes and for evacuation in the actual accelerators.

§ 3. WAVEGUIDES HAVING CONSTANT PHASE VELOCITIES

(i) *The first experimental waveguide*

The theoretical approach was an improved form of that originally given by Cutler (1944). The phase velocity was designed to be equal to the velocity of

light for an air wavelength of 10.00 cm. The inner hole was chosen to give a maximum axial field of 100 kv/cm. for a radio frequency power of 2 mw. (this being the highest power readily available at present). The dimensions were as follows: number of corrugations per wavelength $N=10$, pitch of corrugations $D=1$ cm., ratio of width of corrugation to pitch $K=0.75$, inner radius of corrugations $a=1$ cm., outer radius of corrugations $b=3.944$ cm. Unlike all subsequent guides, the manufacturing tolerances were of the order of ± 0.001 in., although care was taken to prevent any build-up of tolerances over the total length (25 cm.), and the material used was brass, the sections being clamped together without soldering.

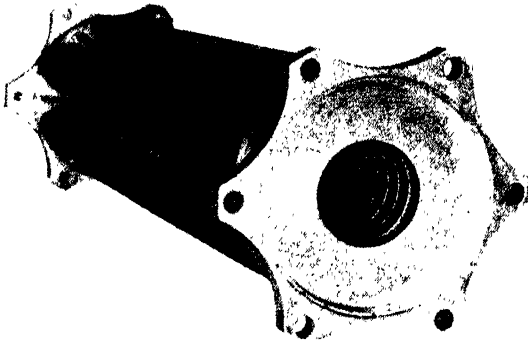


Figure 1. 40 cm. length of corrugated circular guide.

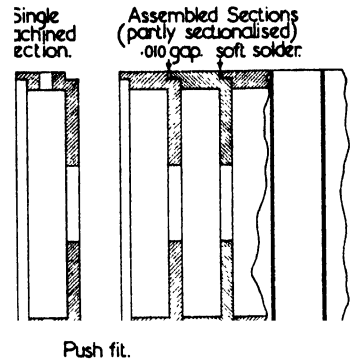


Figure 2. Construction of corrugated circular waveguide. The sections are clamped together and the 0.010 in. gap is filled with soft solder as shown.

Measurements of phase velocity were carried out using a low power klystron oscillator and a coaxial line test bench. The feed into the guide was a coaxial probe, approximately a quarter wavelength long, introduced along the axis. The amount of power fed into the corrugated guide was only a small fraction of that available and could not be improved upon by any simple change of dimensions, but was adequate for this purpose. A tuned crystal detector of standard form (Huxley 1947) fitted with a small pick-up loop was inserted through the measuring holes in the outer wall so as to be lightly coupled to the magnetic field. The phase of the magnetic field in the corrugations varied with distance in the same manner as the phase of the fields on the axis at the centre of each corrugation.

With the end of the guide short-circuited by a flat metal plate the standing wave pattern was plotted in a similar manner to those shown for later cases, e.g. figure 3. The distance between adjacent minimum values gives half the guide wavelength, from which the phase velocity is given by $v_p = c\lambda_0/\lambda$ where λ is free space wavelength, λ_0 is the wavelength in the guide, c is the velocity of light. For $\lambda=10.00$ cm. the guide wavelength was of the order of 10 cm. At shorter air wavelengths the phase velocity decreased as expected and standing wave patterns were readily plotted for guide wavelengths as short as 3 cm. ($v_p=0.3c$). At longer air wavelengths the guide wavelength became indeterminate because of a rapid onset of attenuation. This was later found to be mainly due to the poor frequency spectrum of the oscillator combined with the fact that the low-frequency cut off was less than 10 Mc/s. from the design

frequency (2998 Mc/s). Trouble was also experienced with bad contacts between the unsoldered sections.

(ii) Accurate experimental results

The first guide had indicated that a less frequency sensitive system should be employed in the early stages. A guide was therefore designed on the basis of 25 kV/cm. (for 2 MW. of R.F. power) with dimensions $N=10$, $D=1$ cm., $K=0.75$, $a=2$ cm., $b=4.288$ cm., total length of guide 40 cm.

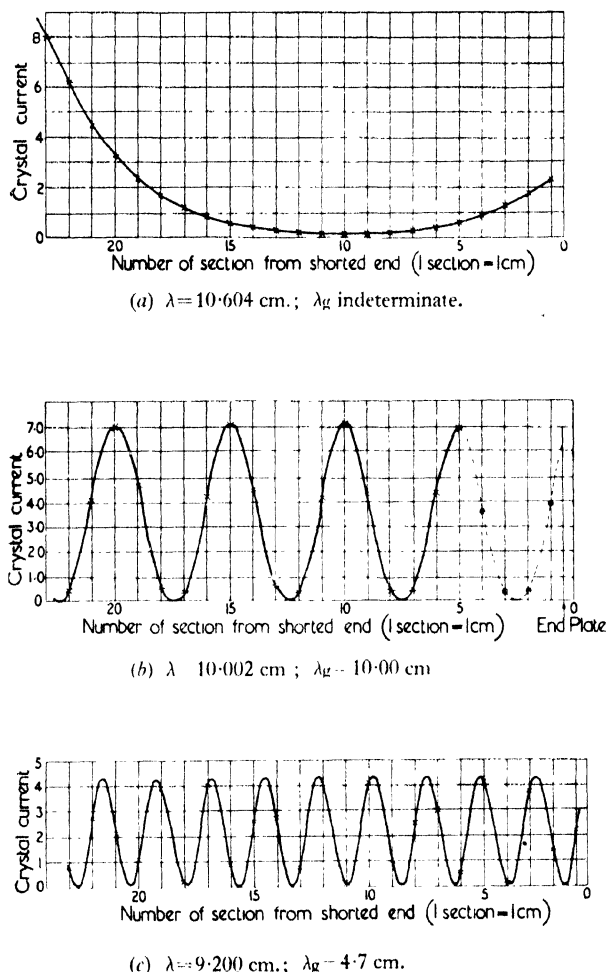


Figure 3. Corrugated circular guide. Variation of standing wave pattern with frequency. $a/\lambda = 0.2$; $K = 0.75$; 1 cm. pitch.

Typical standing wave curves for the short-circuited guide for values of air wavelength around 10.0 cm. are shown in figures 3 (a), (b), (c). Graphs of the relation between the air and guide wavelengths showed fairly close agreement between the experimental results and the theory used to design the guide.

The length of the guide was eventually increased to 100 cm. in order to obtain a reliable measurement of attenuation. The accuracy of the standing wave measurements was again improved and slight irregularities in manufacture and

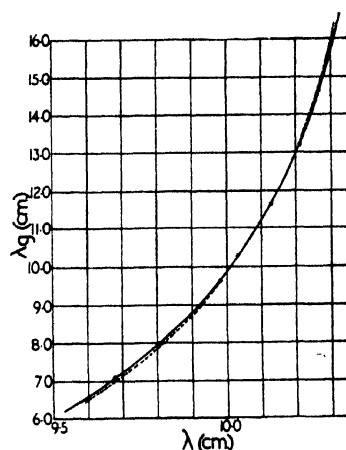
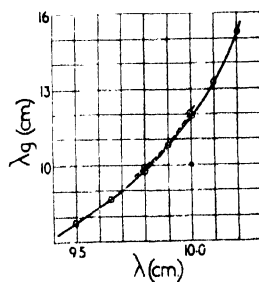


Figure 4. $\lambda - \lambda_g$ corrugated circular guide. (1 metre length.) $a/\lambda = 0.2$; $K = 0.75$; 1 cm. pitch.



X simple theory.
Figure 5. $\lambda - \lambda_g$ for corrugated circular guide with 2 cm. pitch. $K = 0.875$. $a/\lambda = 0.2$. b calculated on simple theory to give $\lambda_g = 10$ cm. for $\lambda = 10$ cm.

—○— experimental
— — — — new theory (Walkinshaw).

assembly were smoothed out. The final curve of λ and λ_g (figure 4) shows remarkable agreement with the more rigorous theoretical treatment (Walkinshaw 1948).

The group velocity can also be obtained from this curve. The general expression is

$$\text{Group velocity } v_g = c(d\lambda/d\lambda_g)(\lambda_g/\lambda)^2;$$

at $\lambda_g = \lambda$ (i.e. the design wavelength), $v_g = c(d\lambda/d\lambda_g)$. The slope of the curve at $\lambda_g = \lambda$ is $d\lambda_g/d\lambda = 10.75$, giving a value for v_g of $0.093 c$, which is again in close agreement with the theoretical value.

An attempt was made to increase still further the accuracy of phase measurement by means of a "phase meter". The apparatus included a "magic T" (Huxley 1947), the properties of which were used for comparing the actual phase of the magnetic field near the outer wall of a corrugation with a small part of the direct power from the R.F. source, whose amplitude and phase could be accurately varied.

The guide was terminated in a long wooden cone, which was adequate for normal standing wave measurements and could not be improved upon. The slight residual standing wave produced a large cyclic variation on the phase pattern along the guide, which was little better than plotting the standing wave pattern.

Accurate measurement of the attenuation was found to be impossible by the more usual waveguide methods. With the end of the 1-metre section short-circuited, the standing wave pattern showed no obvious signs of attenuation—the standing wave ratio at the input end appeared as large as at the short-circuited end. The normal method resorted to in such a case is to measure the shape of the minimum, but, assuming a minimum to occur precisely at a monitoring hole, the next measurable point is 1 cm. away, and the range of detector crystal currents involved is far too great. Instead of moving the detector a method involving slight frequency changes had to be employed.

The procedure was as follows: With an air wavelength of 10.00 cm. a corrugation near the input end was chosen which was as close as possible to a minimum in the standing wave. The input frequency was then slightly changed until an exact minimum was obtained. The frequency was then adjusted to the two points (one on either side) where the crystal current was twice that at the minimum (allowance being made for changes in power level in the corrugated guide by monitoring in the corrugation nearest to the short-circuited end). The guide was sufficiently long for the experiment to be carried out with very small changes in frequency (of the order of 1 Mc/s). A simple calculation gives the standing wave ratio, n , in terms of the total frequency excursion, the length l from the measuring point to the short-circuited end of the guide, and the group velocity v_g already obtained:

$$n = (\pi l / v_g)(\Delta f / f) \quad (n < 1),$$

$$\text{attenuation} = (10/l) \log [(1+n)/(1-n)] \quad \text{db/metre.}$$

The procedure was repeated with a measuring point near the centre of the guide. The difference between the total losses measured in the two cases gave the attenuation for twice the length of guide between the two measuring points, independent of any imperfections in the short-circuiting plate. The average value so obtained was 0.85 db/metre—this being approximately twice theoretical.

This is an encouraging figure since even plain waveguides will give values twice theoretical merely due to imperfections of the metal and its surface. Careful cleaning of the guide gave no appreciable improvement, and the excess loss was therefore assumed to be in the joints between the sections.

(iii) Reduction of the number of corrugations per guide wavelength

It is desirable from the point of view of attenuation to reduce the number of corrugations per guide wavelength, and it has been shown theoretically (Walkinshaw 1948) that optimum efficiency of an accelerator is obtained when N is approximately 4.

A guide was therefore constructed with a corrugation pitch of 2 cm. ($N=5$ at $\lambda=\lambda_g=10$ cm.). The dimensions were: $K=0.875$, $a=2$ cm., $b=4.171$ cm., obtained from early theory, total length 40 cm. The measurement of guide wavelength became even more difficult because of the reduced number of discrete measuring points per guide wavelength. A curve showing the relationship between the guide and air wavelengths is given in figure 5. Although the experimental results (full line) differed considerably from the values given by the early theory used to design the guide (represented by the cross at $\lambda_g=\lambda=10$ cm.) there was extremely close agreement with the final treatment (Walkinshaw 1948) whose validity had not been so obviously demonstrated for guides with 1 cm. pitch ($N=10$).

(iv) General phase velocity measurements

In order to demonstrate the general behaviour of this type of guide, phase velocity measurements were carried out with a set of guides having the characteristics shown in table 1.

Table 1

Guide no.	Pitch (cm.)	K	a (cm.)	b (cm.)*	Guide no.	Pitch (cm.)	K	a (cm.)	b (cm.)*
1	1	0.75	1.0	3.944	4	2	0.875	1.0	3.894
2	1	0.75	1.5	4.085	5	2	0.875	1.5	4.002
3	1	0.75	2.0	4.288	6	2	0.875	2.0	4.171

* b for $\lambda_g = \lambda = 10$ cm. from original theory.

An oscillator of much narrower frequency spectrum than had been available previously was used throughout. The resulting standing wave patterns for the guides with an inner radius as small as 1 cm. were quite good in spite of their high frequency sensitivity.

Figures 6(a) and 6(b) show graphs of λ_g against λ for the guides having 1 cm. and 2 cm. pitches respectively.

(v) Corrugated guides with phase velocities less than that of light

Two guides were constructed having phase velocities of $0.4c$ and $0.87c$ for $\lambda=10$ cm., these corresponding to the input and output ends of a proposed 0.5 mev. accelerator. Both guides had a pitch of 1 cm. and $K=0.75$. At the input end, $a=1.3602$ cm. and $b=4.1814$ cm.; at the output end, $a=1.8948$ cm. and $b=4.2751$ cm. Guide wavelengths of 4 cm. and 8.7 cm. were obtained for $\lambda=10$ cm. in close agreement with the theory.

It was therefore concluded that, for all experimental purposes, an adequate theoretical approach had been established.

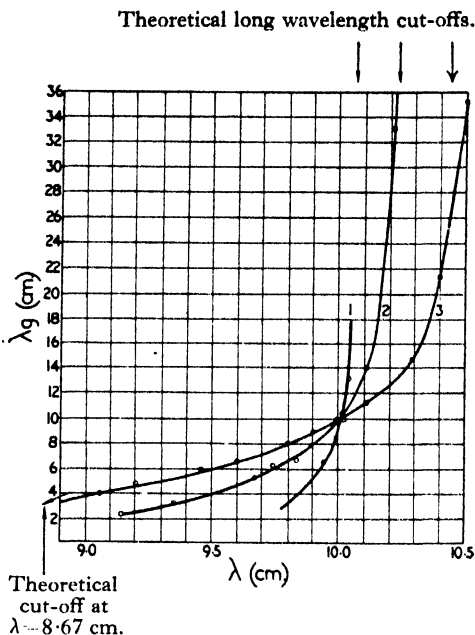


Figure 6 (a). Circular corrugated guides.
Corrugation pitch 1 cm.
Varying a/λ .

Type	a/λ	$2a$
1	0.1	2 cm.
2	0.15	3 cm.
3	0.2	4 cm.

(All guides calculated to give $\lambda_g = 10.0$ cm. at $\lambda = 10.0$ cm.)

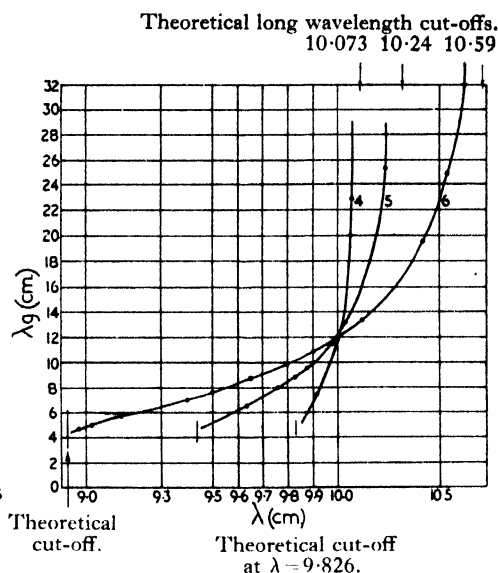


Figure 6 (b). Circular corrugated guides.
Corrugation pitch 2 cm.
Varying a/λ .

Type	a/λ	$2a$
4	0.1	2 cm.
5	0.15	3 cm.
6	0.2	4 cm.

(All guides calculated to give $\lambda_g = 10.0$ cm. at $\lambda = 10.0$ cm.)

§ 4. FEEDS FOR CORRUGATED GUIDES

The design of suitable feeds from the standard rectangular waveguide to the circular corrugated guide presented many difficulties. The primary consideration is that the feed should allow the electrons to be injected along the axis of the corrugated guide, and in doing so should not expose the beam to appreciable fields until it enters the accelerating field proper.

The simple quarter wavelength probe along the axis was very soon discarded for anything except low power measurements.

The obvious method of slowly tapering from a smooth E_{01} guide to the corrugated guide is very clumsy; the beam is subjected to undesirable fields, and its only virtue, that of wide bandwidth, is quite unnecessary for the application considered.

(i) Annular hole feed

Several mechanically suitable feeds were tried, one of which yielded quick results since some theoretical design was possible. It consisted of a coaxial line tapered to a low impedance and fed directly into the first corrugation as shown in figure 7. The electron beam could enter through a hole down the centre of the inner conductor and R.F. power was fed to the coaxial line from a standard rectangular waveguide ($3 \times 1\frac{1}{2}$ in. o.d.) by means of a "door knob" transformer (Huxley 1947).

The conditions which it was essential, or desirable, to satisfy were: (a) the coaxial line should be of a convenient impedance (40 ohms); (b) the coaxial line to be below cut-off for the first waveguide mode (H_{11}) so as to eliminate any possibility of propagation of an asymmetrical mode from the rectangular guide into the corrugated guide; (c) the inner conductor to be as large as possible when satisfying the above conditions; (d) the coaxial line to be tapered to match into the first corrugation at a point between its inner and outer radius; (e) the electric fields should be limited to 100 kv/cm. or less for an R.F. power of 2 MW. to reduce the possibility of breakdown.

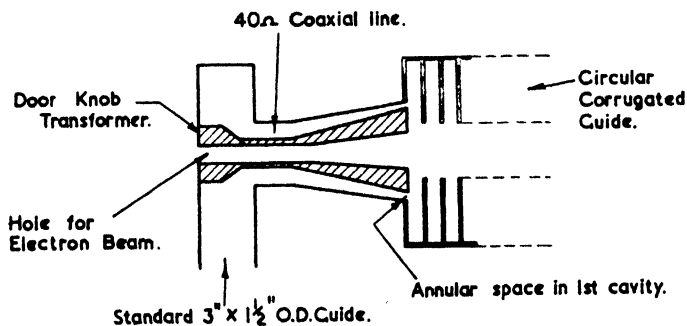


Figure 7. Arrangement of annular hole feed.

Since the corrugated guide is effectively a high Q device we expect the value of the field strength at any point to be independent of the point at which the wave from the source is introduced.

In order to obtain a match the wave impedance looking into the annular space must be identical with that of a matched coaxial line with the same inner and outer radii, i.e.

$$E'_0/H'_q = E''_0/H''_q, \quad \dots\dots(1)$$

where E'_0 is a radial component of electric field assumed to have been introduced into the corrugation by the annular gap, H'_q is the magnetic field in the corrugation at the mean radius of the annular gap, E''_0 and H''_q are the fields in a coaxial line. If the coaxial line is terminated everywhere by its own impedance all the power will be transmitted through the annulus, and integrating the Poynting vector over the annular space it follows that

$$\int_s E'_0 H''_q ds = \int_s E''_0 H''_q ds. \quad \dots\dots(2)$$

If it is assumed that the field variations between the inner and outer radii of the annulus are small for both the coaxial line and the corrugation, then the integration signs can be removed. Hence if H'_0 and H''_q are equated at the mean radius of the annulus, the equality of E'_0 and E''_0 follows automatically.

The 40-ohm coaxial line chosen to satisfy the required conditions had inner radius $r=1.03$ cm. (0.406 in.) and outer radius $R=2$ cm. (0.787 in.). This allowed a fairly large diameter hole (5/8 in. maximum) for the electron beam. The fields in a coaxial line (in M.K.S. units) are:

$$E''_0 = A'(1/\rho), \quad \dots\dots(3)$$

$$H''_q = A'(\epsilon/\mu)^{1/2}(1/\rho). \quad \dots\dots(4)$$

The constant A' is found by integrating the Poynting vector over the cross-sectional area normal to the axis of propagation:

Mean power

$$\bar{W} = \frac{1}{2} \int_s E_\phi'' H_\phi'' ds, \quad \dots\dots(5)$$

Hence

$$A' = \left[\frac{\bar{W}}{\pi} \left(\frac{\epsilon}{\mu} \right)^{\frac{1}{2}} \frac{1}{\ln R/r} \right]^{\frac{1}{2}} \quad \dots\dots(6)$$

and

$$H_\phi'' = \frac{1}{\pi\rho} \left[\frac{\bar{W}}{120} \cdot \frac{1}{2.303 \log R/r} \right]^{\frac{1}{2}} \text{ amp/cm.} \quad \dots\dots(7)$$

In the centre portion of the corrugated guide by a similar process, in the limit as $v_p \rightarrow c$ (i.e. for a guide with phase velocity = c):

$$H_\phi = (\rho/\pi r^2)(\bar{W}/30)^{\frac{1}{2}} \text{ amp/cm.} \quad \dots\dots(8)$$

Inside the corrugations the magnetic field is given by

$$H_\phi' = \frac{iB}{120\pi} \left[J_1(2\pi\rho/\lambda) - \frac{J_0}{Y_0} \frac{2\pi b/\lambda}{2\pi b/\lambda} Y_1(2\pi\rho/\lambda) \right]. \quad \dots\dots(9)$$

The constant B can be found by equating H_ϕ to H_ϕ' at $\rho = a$, and hence H_ϕ' and H_ϕ'' can be equated for a range of values of ρ . As ρ increases the gap width becomes very small, an undesirable feature since the inner of the coaxial line is liable to sag.

For the guide having $a/\lambda = 0.2$ and a pitch of 1 cm. the value of ρ was chosen to be 2.5 cm. so that the feed would be clear of higher order modes near the mouth of the corrugation. The gap width was then 0.218 cm. and the electric field E_ϕ'' across it 21.9 kV/cm. (for 2 MW. R.F. power). An experimental feed made up precisely as calculated was tested, and found to have a standing wave ratio of 0.56/1 which varied very little with changes in frequency.

The effect on the standing wave ratio of eccentricity of the inner portion of the feed with respect to the outer, and of axial displacement of the inner, was extremely small, but the introduction of the hole for the electron beam had a marked effect. An optimum of the standing wave ratio (0.7/1) was obtained for a hole diameter of 1.10 in., probably due to the conditions in the end cavity of the corrugated guide approximating more exactly to the theoretical case.

No high power measurements were made at this stage but it seemed clear that no breakdown troubles would occur. The annular hole appeared to be the limiting factor, but even here the calculated fields showed a wide margin of safety. That these assumptions were true was borne out by later experience with an actual accelerator.

A similar feed was designed and constructed for the very critical guide with $a/\lambda = 0.1$. The mean radius chosen was 1.50 cm. giving a gap width of only 0.058 cm. Even so the feed showed signs of working, its standing wave ratios being 0.53/1, 0.52/1 and 0.29/1 at air wavelengths of 9.90, 9.93 and 9.97 cm. respectively.

The measure of success obtained led to this type of feed being adopted for the first linear accelerator. The fields in the corrugated guide are more complicated when the guide has a phase velocity less than c . Then $1 - (c/v_p)^2$ becomes negative.

Hence the magnetic field within the inner diameter of the corrugations is given by

$$H_{\varphi} = E_0 \left[\frac{\mu}{\epsilon} \left\{ \left(\frac{c}{v_p} \right)^2 - 1 \right\} \right]^{-\frac{1}{2}} I_1 \left[\frac{2\pi\rho}{\lambda} \left\{ \left(\frac{c}{v_p} \right)^2 - 1 \right\}^{\frac{1}{2}} \right], \quad \dots\dots (10)$$

where E_0 is the accelerating field along the axis. The calculation then proceeds as before.

(ii) Short door-knob feed

One other feed has been found to work experimentally. It offers several advantages over the annular hole feed. The length of coaxial line, which was difficult to assemble correctly and greatly obstructed the electron beam, has been effectively removed.

The object was to feed the first corrugation by attaching it directly and at right angles to the broad face of the rectangular guide, using for convenience a hole of the same size as the inner diameter of the corrugated guide. The coaxial line and door-knob transformer were then compressed into a short door-knob forming a circular projection of appropriate dimensions (as shown in figure 8) from the outer broad face of the rectangular guide to the first hole of the corrugated guide. For almost any diameter of post, P , with its corresponding radius of curvature S , it is possible to obtain a good match by adjustment of its

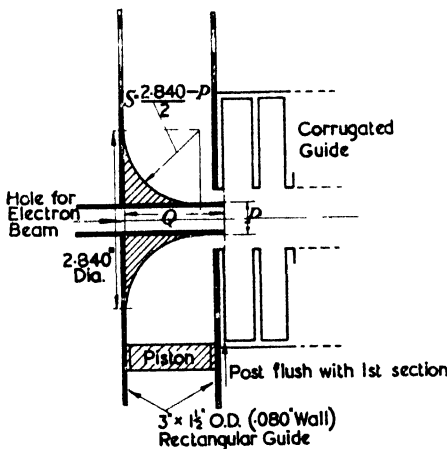


Figure 8. Short "door knob" feed.

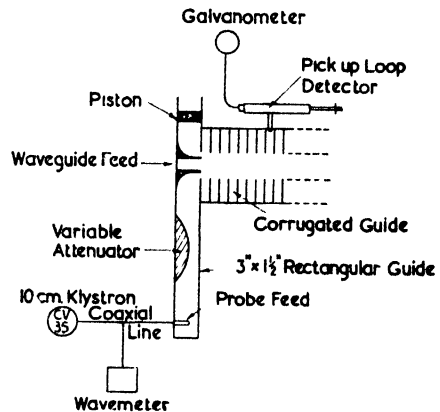


Figure 9. Schematic diagram of improved test gear using short "door knob" feed.

length Q , using the terminating piston in the rectangular guide as another variable. It is then found that with small diameters the post does not reach the entrance to the first cavity, with large diameters the post projects into the cavity, and as P is further increased the optimum position no longer gives a perfect match. It is possible to choose a diameter which will give a perfect match with the post flush with the first section.

This feed has been used for uniform guides with values of a/λ as small as 0.1 although the setting of the variables to obtain a perfect match is rather critical. It is also being incorporated in a 4 mev. accelerator, and although no high power tests have yet been carried out, no breakdown troubles are expected.

The test equipment and procedure is very much simplified when a purely waveguide feed of this type is used, as shown in figure 9. The mechanical arrangement allows the door-knobs to be interchanged with great ease.

§ 5. CORRUGATED CIRCULAR WAVEGUIDE WITH CONTINUOUS VARIATION OF PHASE VELOCITY

A theoretical treatment for guides of uniform phase velocity being established, the design of a waveguide system for an actual accelerator followed similar lines. The velocity of the injected electrons was chosen to be $0.4c$, equivalent to a gun voltage of 45 kv. In order to simplify the manufacture a constant corrugation pitch of 1 cm. was chosen. The number of corrugations per guide wavelength at the beginning of the accelerator was therefore considerably less than ten, but the theory had been shown to be accurate down to five corrugations. In order to simplify the computation of electron energy against distance down the accelerator, the inner diameter of the corrugations was reduced at low phase velocities in order to maintain a constant accelerating field on the axis.

Two such systems have been designed (Walkinshaw 1948) and manufactured.

(i) 0.5 Mev. *bunching section of a long accelerator*

It was originally considered that the initial section of a linear accelerator working from 45 kev. injected electrons up to the order of 0.5 Mev. would present the major difficulties, since we have large variations of phase velocity, strong bunching forces, and also strong defocusing forces. The first objective was therefore a short length (40 cm.) of an accelerator giving 540 kev. for 1 MW. R.F. power capable of becoming efficient when extended to a length of some 10 metres.

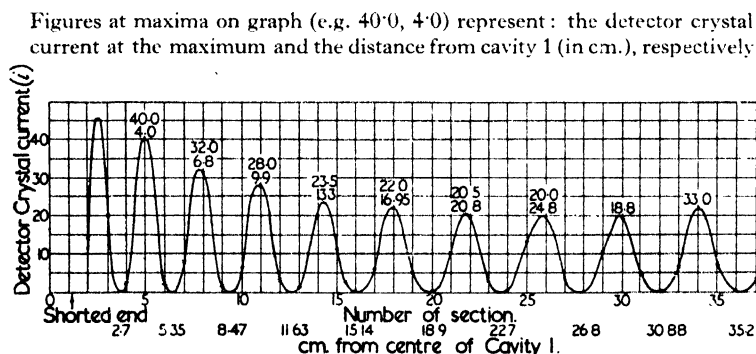


Figure 10. Standing wave pattern in 40 cm. tapered guide for 0.5 Mev. linear electron accelerator. Voltage maxima are obtained from $\sqrt{i}$ (crystal current) $\times$ (factor taking into account changes in wall thickness).

(a) *Corrugated guide.* The phase along the guide was again obtained from a plot of the standing wave pattern with one end short-circuited. The maxima and minima determine phase intervals of $\pi/2$. The theoretical and experimental phase curves are shown in figure 10. The agreement is a remarkable tribute to the theory, manufacture and measurements. It can be seen from the standing wave pattern that the magnetic field which is measured near the outer wall of the corrugations decreases from the $0.4c$ end. This is not due to attenuation. At $v_p < c$, both the magnetic field H_ϕ and electric field E_z varying, as I_1 and I_0

functions of ρ (Bessel functions of an imaginary argument), have minimum values on the axis. The inner hole having therefore been reduced to give uniform accelerating field on the axis, the fields at the outside of the guide will be greatest at the low velocity end. The maxima in the standing wave pattern represent in the ideal case the squares of the magnetic fields (H_ϕ) at the outer wall. By working at very low crystal currents and applying corrections for crystal law, and certain mechanical factors, H_ϕ was obtained with some measure of accuracy, and compared graphically with calculated values. The agreement was close, considering the complexity of the measurements, and was a convincing demonstration of the uniformity of the axial accelerating field.

(b) *Input and output R.F. feeds.* Feeds of the tapered coaxial line type were chosen for this accelerator. The dimensions were finally obtained from a consideration of sets of graphs such as that shown in figure 12 for the output end, and are given in table 2.

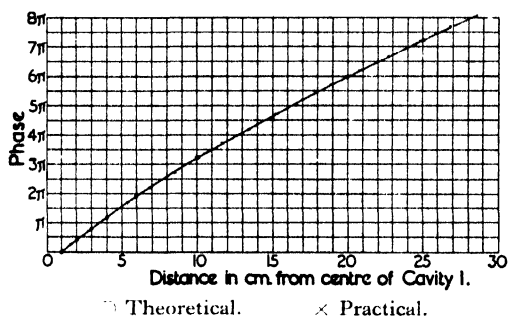


Figure 11. Phase in 40 cm. tapered guide.

Next points: 28.8, 8.2π ; 30.88, 8.7π ; 33.0, 9.2π ;
35.2, 9.7π .

From voltage max. and min. curve (figure 10).
Adjusted to same phase at 2nd maximum (4.0).

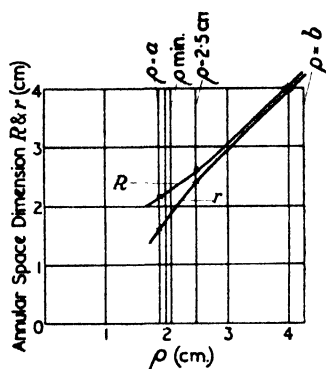


Figure 12.

Design data for output feed.
(0.5 MeV. accelerator.)

Table 2

	ρ (cm.)	a (cm.)	b (cm.)	Z_0 (ohms)
Input feed	1.8	1.725	1.875	5.0
Output feed	2.5	2.409	2.591	4.4

Experimental feeds were made to these dimensions and the size of the electron hole was adjusted to give the best match. The sharp edges of the annular gap and the electron hole were slightly rounded. The standing wave voltage ratios were then 0.55/1 for the input feed and 0.53/1 for the output feed. It was found impracticable, by alteration of those dimensions easy to vary, to obtain a better match for the feeds alone. However, it was again found possible, when the feeds combined with their waveguide to coaxial transformers were assembled in the end sections of the accelerator envelope, to obtain an improved match by axially adjusting the inner portions of the feeds, thereby obtaining standing wave ratios of 0.58/1 and 0.76/1 for the input and output feeds respectively. The final matching was carried out by the use of inductive irises in the main rectangular guide as near to the waveguide to coaxial transformers as possible, and gave standing wave ratios of 0.976/1 for the input end and 0.94/1 for the output end. The standing wave ratio of the corrugated waveguide system

complete with feeds was found to be 0.895/1 looking into the input (0.4 c) end and 0.94/1 looking into the output (0.87 c) end, the discrepancy being attributed to the attenuation of the corrugated guide.

When finally assembled as an accelerator (Fry, Harvie, Mullett and Walkinshaw 1947) with a vacuum-tight steel dummy load and a directive feed for monitoring, the overall standing wave ratio presented to the magnetron was found to be better than 0.95/1 at a wavelength of 10.00 cm.

(ii) 4 Mev. section of a long accelerator

The second experimental waveguide system was again designed as the initial section (2 metres) of an accelerator capable of being extended to some 10 metres in length. The peak accelerating field per megawatt of R.F. power was chosen to be the same as for the 0.5 Mev. accelerator, but the power required to give an initial stable electron position of 45° was increased to 2 MW. The first 40 cm. length was designed so that the R.F. phase velocity was slightly greater than the velocity of the stable electron, in order to decrease the phase angle linearly from 45° to 35° , and hence to increase the field experienced by the stable electron as the amplitude of the phase oscillations within the bunch decreased. At 40 cm. the phase velocity is slightly reduced to that of the stable electron so as to maintain a constant phase angle of 35° for the remainder of the 2 metres.

The corrugation pitch was again chosen to be 1 cm. throughout, and the construction was almost identical with that of the corrugated guides for the 0.5 Mev. accelerator.

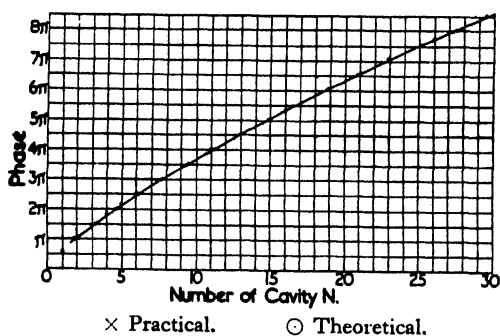


Figure 13 (a). Phase at beginning of 2 metre tapered guide (for 4 Mev. linear electron accelerator).

Relatively adjusted to coincide at 7π ($N=23$) since this is an extremely accurate point.

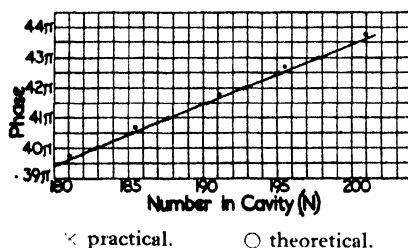


Figure 13 (b). Phase at end of 2 metre tapered guide.

In order to correct the phase the wavelength should be decreased, i.e. effectively deeper corrugations. This could not be achieved with metallic screws.

(a) *Phase measurements.* A standing wave curve was plotted for the complete 2 metre length and a curve of phase against length obtained. Figures 13 (a) and (b) show the interesting regions near the beginning and end of the accelerator. The theoretical phase curve has been adjusted to fit the experimental curve at cavity 23, this being an accurately defined experimental point. Slight fluctuations in phase in the cavities before this are relatively unimportant since the electron velocity will be well below c and can adjust itself to the wave. The error between theoretical and experimental phase at the end of the 2 metres is 18° (in 22 wavelengths), corresponding to a mean wavelength error of 1 part in 400. This error is cumulative throughout rather than random, hence it would appear that errors due to

manufacturing tolerances are negligible and that the purely theoretical error is far smaller than had been expected. In the actual accelerator such an error could be corrected by a small change in frequency. There was also some indication on the graph of the phase jump in the region of the 40th section.

The 18° phase error in 22 wavelengths is tolerable for this accelerator. Longer accelerators would require some correction to be applied at intervals of this order—or possibly less (Harvie 1948). The introduction of 3/16 in. diameter dielectric or metal rods into the corrugations has been tried, and although a phase correction of the required order could be produced with only a few rods, it was necessarily in one direction only and was accompanied by a sharp rise in attenuation (possibly due to the setting up of an asymmetrical mode) especially in the case of the dielectrics. Short correcting sections of corrugated guide provide the best solution.

(b) *R.F. feeds.* The short door-knob feed was used for this accelerator. For the 0.4 c end of the accelerator the inner diameter of the corrugated guide was 2.72 cm.; the largest diameter of post which would give a perfect match when flush with the first section was 1.7 cm. This gave a 1.5 cm. diameter hole for the injection of electrons from the gun.

At the output end of the accelerator it is permissible for the post to project into the last section; the post diameter was therefore increased to the maximum of 3.2 cm. with a projection into the last cavity of approximately 3 mm. in order to obtain a perfect match. This gives an electron hole of 3 cm. diameter and a prospect of higher beam current leaving the accelerator. If the hole into the corrugated guide is increased in order to fit a still larger diameter of post, it is difficult to obtain a good match since the projection into the corrugated guide becomes excessive. The rectangular piston had to be soldered into the guide to obtain good contact and also to form a vacuum joint, hence it was found desirable to have a preset screw projecting through the centre of the piston and acting as a very fine variable. In this way final standing wave ratios of 0.96/1 to 0.97/1 were obtained for the input and output feeds.

The behaviour of these feeds to changes in frequency is of considerable interest. It is necessary that a good match should be maintained over the working band of frequencies. If an arbitrary figure of $\pm \frac{1}{3}$ Mc/s. for the magnetron frequency stability is chosen as a basis for design, then this particular accelerator could be extended to a length of some 10 metres, and the associated components (in particular the feeds here described) need only be efficiently matched over a bandwidth of this order.

The bandwidth of the present feeds for a standing wave ratio of 0.9/1 or better are for the input ± 3 Mc/s., for the output ± 4 Mc/s. The BM 735 magnetron, which will again be used, can be pulled in frequency by the order of ± 3 Mc/s. It will therefore be possible even with this two metre length to change the phase at the end of the accelerator by at least $\pm 0.4\pi$ without deterioration of the R. F. system. This will adequately cover the interesting regions of operating conditions.

(c) *Overall performance.* The standing wave ratio of the complete corrugated guide and feeds was 0.96/1.

The attenuation of the complete system was determined by replacing the dummy load with a short-circuited guide and measuring the standing wave ratio in the input rectangular guide. The mean value obtained was 2.5 db. By removing 40 cm. sections the guide itself was found to have a loss of approximately 0.9 db. per metre, leaving a residue of about 0.7 db. for the two feeds and various

joints in the rectangular guides. This attenuation is again twice the theoretical value but is not excessive for this accelerator. It will, however, be desirable in future accelerators of efficient lengths to increase the pitch to 2 cm., except in the initial region, so as to halve the losses. It will also be acceptable (although not imperative) if new methods of construction now in hand give losses near to the theoretical values.

ACKNOWLEDGMENTS

The authors wish to express their appreciation of the assistance given to them by their colleagues during the carrying out of this work in the Electronics Group of the Atomic Energy Research Establishment at the Telecommunications Research Establishment, Malvern. Acknowledgment is made to the Director, Atomic Energy Research Establishment, Ministry of Supply, for permission to publish these results.

REFERENCES

- CUTLER, C. C., 1944, *Bell Tel. Labs. Report*, MM/44/160/218.
 FRY, D. W., HARVIE, R. B. R-S-, MULLETT, L. B., and WALKINSHAW, W., 1947, *Nature, Lond.*, **160**, 351.
 HARVIE, R. B. R-S-, 1948, *Proc. Phys. Soc.*, **61**, 255.
 HUXLEY, L. G. H., 1947, *A Survey of the Principles and Practice of Waveguides* (Cambridge: The University Press), pp. 80, 188, 95 (in order of reference in text).
 WALKINSHAW, W., 1948, *Proc. Phys. Soc.* **61**, 246.

Effect of Azimuthal Inhomogeneities in the Magnetic Field of a Betatron or Synchrotron

By F. K. GOWARD

Telecommunications Research Establishment, Malvern

MS. received 14 February 1948

ABSTRACT. Azimuthal inhomogeneities in the magnetic field of a betatron or synchrotron cause a forced oscillation of the electron; this forced oscillation may be analysed into oscillations whose frequencies are simple multiples of the rotational frequency of the electron. The complete motion of the electron may be described by adding to the forced oscillation the normal free oscillation of the electron, the period of which is not, in general, commensurate with the rotational frequency.

In this paper are considered quantitatively some particular examples of the types of electron oscillation which may be encountered, with special emphasis on the problems of injecting an electron into an inhomogeneous field, and of extracting an electron by an applied inhomogeneity.

§ 1. INTRODUCTION

THE effects on the electron orbits of azimuthal inhomogeneities in the magnetic field of a betatron or synchrotron have been considered by several workers (Bohm and Foldy 1946, Clark *et al.* 1946, Goward and Dain 1947); no simple general explanation of the effects and their occurrence and application has, however, been given. It is hoped that the description given here, and the illustrative examples considered, will emphasize the practical importance of these inhomogeneities. There is, admittedly, some doubt as to the correctness of a simple theory of electron orbits in a betatron type field.

(Kerst 1941, Kerst and Koch 1947), but discrepancies between theory and practice can only be attributed to physical conditions extra to those considered here.

§ 2. MOTION OF ELECTRON IN INHOMOGENEOUS FIELD

The undisturbed axial field of a betatron or synchrotron normally follows the radial law

$$H_z = H_0(r/r_0)^{-n}, \quad \dots\dots(1)$$

where H_0 is the field at some particular radius, r_0 , and $0 < n < 1$. An electron with velocity v_0 , where

$$v_0 = (e/mc)H_0r_0 \quad \dots\dots(2)$$

would describe an orbit at r_0 , if injected tangentially at r_0 .

Consider that a small axial disturbing field is added, its magnitude varying sinusoidally with θ , i.e. the total field is

$$H_z = H_0(r/r_0)^{-n}(1 + \delta_m \sin m\theta), \quad \dots\dots(3)$$

where $\delta_m \ll 1$, and m is any integer. The equation of radial motion in this field, for an electron of velocity v_0 , is

$$d^2r/dt^2 = v_0^2/r - (e/mc)v_0H. \quad \dots\dots(4)$$

Restricting to small displacements, x , from r_0 , where $r = r_0 + x$, equation (4) becomes, using equations (2) and (3) and neglecting terms in x^2 etc.,

$$d^2x/dt^2 = (v_0^2/r_0)[(1 - x/r_0 \dots\dots) - (1 - nx/r_0 \dots\dots)(1 + \delta_m \sin m\theta)]. \quad \dots\dots(5)$$

This is conveniently put in terms of θ by the substitutions $d\theta/dt = \omega_0$ and $\omega_0 = v_0/r_0$ giving

$$d^2x/d\theta^2 = r_0[-x/r_0 + nx/r_0 - \delta_m \sin m\theta + (nx/r_0)\delta_m \sin m\theta].$$

The last term may be neglected, since it is a product of two small quantities, giving

$$d^2x/d\theta^2 + (1 - n)x = -r_0\delta_m \sin m\theta. \quad \dots\dots(6)$$

The general solution of this equation is

$$x = [r_0\delta_m \sin m\theta/(m^2 + n - 1)] + [A \sin (1 - n)^{\frac{1}{2}}\theta + B \cos (1 - n)^{\frac{1}{2}}\theta], \quad \dots\dots(7)$$

the first term representing the forced oscillation, and the second term the free oscillation, A and B being arbitrary constants determined by the initial conditions of the electron. In injection problems the two terms may be comparable, but in extraction problems there is no free oscillation initially and the disturbing field is normally introduced slowly, compared with the oscillation period, so that no free oscillations are excited.

Any arbitrary disturbing field may be synthesized by a Fourier series of terms corresponding to equation (3), i.e.

$$H_z = \dot{H}_0(r/r_0)^{-n} \left[1 + \sum_{m=0}^{\infty} \delta_m \sin (m\theta + \alpha_m) \right], \quad \dots\dots(8)$$

where α_m allows for azimuthal shift of the disturbing field. This gives, for the electron motion,

$$x = \left[r_0 \sum_{m=0}^{\infty} \epsilon_m \sin (m\theta + \alpha_m) \right] + [A \sin (1 - n)^{\frac{1}{2}}\theta + B \cos (1 - n)^{\frac{1}{2}}\theta], \quad \dots\dots(9)$$

where $\epsilon_m = \delta_m/(m^2 + n - 1). \quad \dots\dots(10)$

The terms of equation (10) usually become negligible as the order of the harmonic, m , becomes large.

To illustrate the above analysis, some examples of practical importance will now be studied.

§ 3. INJECTION INTO A BETATRON

The effect of eddy-current disturbing fields on injection into a betatron may conveniently be studied by taking an example, the original Kerst-type 2.3 or 4 mev. betatron. The magnet construction in this machine leads naturally to out-of-phase fields which are of significant magnitude at injection, and these fields vary in a conveniently regular manner with azimuth.

Some measurements on these fields for a 4 mev. betatron, whose radial field varies according to the law $H = H_0(r/r_0)^{-0.7}$, where $r_0 = 7.5$ cm., the equilibrium orbit radius, are shown in figure 1. The curve is arbitrarily arranged so that its mean ordinate is zero; this is permissible, since a mean value different from zero merely entails injecting the electrons at a time slightly different from that with no eddy-currents present. Taking an injection voltage of 2 kv., which corresponds to a mean field of 19.9 gauss at 7.5 cm. radius, harmonic analysis of the full curve of figure 1 gives

$$H_z = H_0(r/r_0)^{-n} \left[1 + \sum_{m=1}^{\infty} \delta_m \cos m\theta \right], \quad \dots (11)$$

where $\delta_1 = -0.017$, $\delta_2 = -0.120$, $\delta_3 = 0.005$, $\delta_4 = 0.028$ etc. The forced oscillation therefore has the coefficients $\epsilon_1 = -0.024$, $\epsilon_2 = -0.032$ and all other coefficients are of the order of 0.001 or less, and may be neglected. Thus the equation of the forced oscillation is

$$x = -0.18[\cos \theta + (4/3) \cos 2\theta] \text{ cm.} \quad \dots (12)$$

This oscillation is shown in figure 2, where it may be noted how the lower harmonics are accentuated: this is an important point to notice in practical betatron design, as has been observed experimentally by Adams, Kerst and Scag (1947).

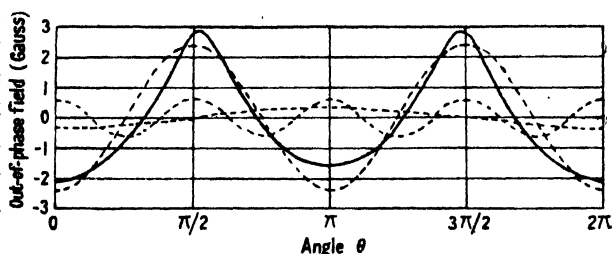


Figure 1. Measured field at injection (full curve), and its harmonic components (dotted curves).

NOTE: Zero out-of-phase field is taken as the mean height of the full curve. For 2 kv. injection voltage and $r_0 \times 7.5$ cm. this zero corresponds to 19.9 gauss main field.

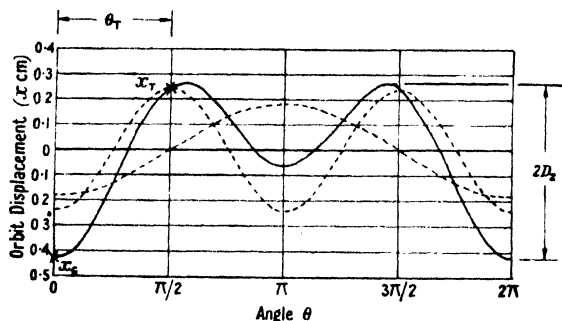


Figure 2. Forced oscillation of orbit (full curve), and its harmonic components (dotted curves). The oscillation corresponds to the field disturbance of figure 1.

In the injection process the forced oscillation of equation (12) combines with a free oscillation, the free oscillation amplitude being determined by the condition that it must combine with the forced oscillation to give an electron coincident with the gun initially. A typical oscillation is shown by the full line of figure 3(a), the oscillation being a combination of the forced oscillation of overall amplitude 0.7 cm. and a free oscillation of overall amplitude 2.9 cm. The injection gun is shown at $r = 8.5$ cm. by the points $G_1, G_2, \dots$ etc., and the target at $r = 4.7$ cm. by the points $T_1, T_2, \dots$ etc.; the target is at right angles to the injection gun. A pure free oscillation of the same amplitude as the free oscillation component of figure 3(a) is shown in figure 3(b), i.e. the oscillation is of 2.9 cm. overall amplitude about a mean radius of 7.05 cm. The figures drawn do not assume damping, since damping will be considered separately later. For this undamped case it may be seen that the target in figure 3(a)

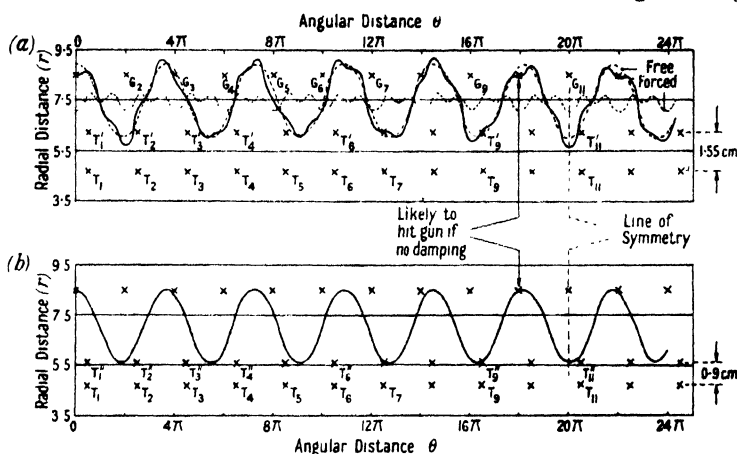


Figure 3. Oscillations in betatron. (a) Forced + free oscillations. (b) Free oscillation.

can be moved 1.55 cm. from its nominal radial position to $T_1', T_2', \dots$ etc., without intercepting the electron at T_7' or T_9' . In figure 3(b), however, a movement of 0.9 cm. only is permissible, to $T_1'', T_2'', \dots$ etc. Thus, with the forced oscillation present, the target may be 0.65 cm. closer to the gun, radially, for the same amplitude of free oscillation. This 0.65 cm. is clearly the radial distance $x_T - x_G$ in figure 2.

Considering now the maximum permissible oscillations with the gun and target fixed at G_1 and T_1 , the amplitudes of the free and forced oscillations are again fixed by the conditions that the gun and target must not be hit and that the electron must be coincident with the gun initially. A free oscillation is permissible, therefore, of overall amplitude approximately 0.65 cm. greater than that permissible with no forced oscillation present, i.e. of 4.45 cm. overall amplitude. This figure is approximate because there is a slight change in the amplitude of the forced oscillation dependent on the mean radius about which the oscillation occurs (see equation (9)); this slight change may be neglected.

A general statement can be made for this particular type of betatron, with an external gun and an internal target at fixed radial distances; if the target is at an angle θ_T relative to the gun (see figure 2), then a free oscillation is permissible, on injection, of amplitude A_2 , where

$$A_2 = A_1 + \frac{1}{2}(x_T - x_G); \quad \dots\dots(13)$$

A_1 is the maximum permissible free oscillation amplitude with no forced oscillation present and is therefore half the radial separation between the gun and the target; $x_T - x_G$ is defined in figure 2. The largest effect occurs if the target is placed at an angle θ_T such that $|x_T - x_G|$ has its maximum value, $2D_2$. Then the limiting values of A_2 are given by

$$A_2 = A_1 \pm D_2, \quad \dots\dots(14)$$

where A_2 takes its maximum or minimum value as the phase of the free oscillation changes by π relative to the gun. This effect of the phase of the forced oscillation may be seen in the initial states shown in figure 5, which is discussed more fully later. The presence or absence of the forced oscillation does not in any way affect the number of revolutions before the electron hits the gun in this undamped case. The only factor which will affect this number is the possibility of the electron hitting the "donut" at some point. This follows from the fact that the full curve in figure 3 (a) passes below the level of 'T₁'. It is possible, however, that the relative magnitudes of the oscillation components, as expressed by equation (13), may have a significant effect on the injection process by changing the damping. This problem must now be studied.

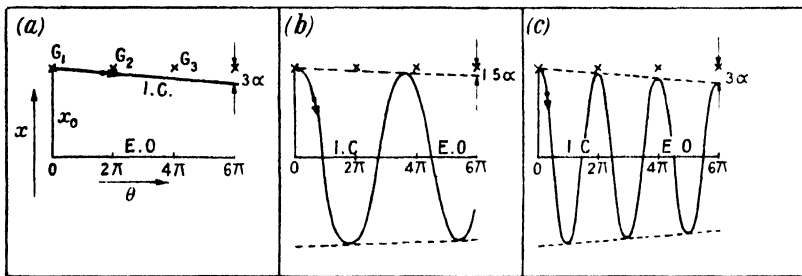


Figure 4. Types of damping.

(a) Shrinkage of instantaneous circle: $x = x_0(1 - a\theta/2\pi)$.

(b) Damping of free oscillation: $x = x_0(1 - a\theta/4\pi) \cos(1 - n)\theta$.

(c) Damping of forced oscillation: $x = x_0(1 - a\theta/2\pi) \cos m\theta$.

NOTE: I.C. denotes instantaneous circle; E.O. denotes equilibrium orbit.

Theories of the mechanism of injection which have been advanced (Kerst and Serber 1941) rely on damping to miss the gun, which is considered to be substantially a point. These theories are not in accord with experiment, failing in particular to explain the approximately linear variation of beam current with injection voltage, and the possibility of injecting with a gun which requires an inordinate damping to miss it. Nevertheless, it seems reasonable to assume that damping is important as a contributory factor in injection and to examine the effect on the damping of the presence of the forced oscillations. Kerst and Serber state that two forms of damping are possible:

(a) An electron may be injected tangentially at a field value and energy such that its radius of curvature is equal to the radius at which the gun is placed. The orbit then shrinks, as shown in figure 4(a), the shrinkage being such that

$$\Delta x/x = -\Delta H/H. \quad \dots\dots(15)$$

From this it follows, since the field H rises linearly with time during injection, that

$$x = x_0(1 - \alpha\theta/2\pi), \quad \dots\dots(16)$$

α is the damping factor, and the relation holds when x/x_0 is near unity.

(b) An electron may be injected at a field value and energy such that its radius of curvature would be r_0 , were it injected tangentially at radius r_0 . A free oscillation is then obtained, as shown in figure 4(b), the damping of the oscillation being half that of equation (16).

To these forms of damping, which may be called, respectively, shrinkage of the "instantaneous circle" towards the equilibrium orbit and damping of the free oscillations about the instantaneous circle, must now be added a further type of damping, the damping of the forced oscillation. The disturbing eddy current field producing the forced oscillation remains constant with time, whereas the main field rises linearly with time. Thus δ_m , the relative amplitude of the disturbing field, varies inversely with time, and ϵ_m does likewise by equation (10). The oscillation amplitude therefore varies inversely with field, i.e.

$$|x| = k/H, \quad \dots\dots(17)$$

which gives a relation of the form of equation (15). Thus the damping of the forced oscillation is the same as the shrinkage of the instantaneous circle and is double the damping of the free oscillation. The expression for damping assumes that δ_m changes inappreciably during one revolution, so that no free oscillations are excited during the change.

Returning now to the example which has been considered for the undamped case, the dampings may be compared with and without forced oscillations present; the argument will be restricted to oscillations which are as large as the space between the gun and target will permit.

The maximum permissible free oscillation, with no forced oscillation present, has an amplitude of 1.9 cm. about an instantaneous circle of 6.6 cm.; this is shown in the left-hand portion of figure 5(a). After r revolutions the amplitude of the oscillation is damped by $0.95\alpha r$ cm. It will be convenient to consider always that $r=9$, since this corresponds to the point G_{10} when the electron is likely

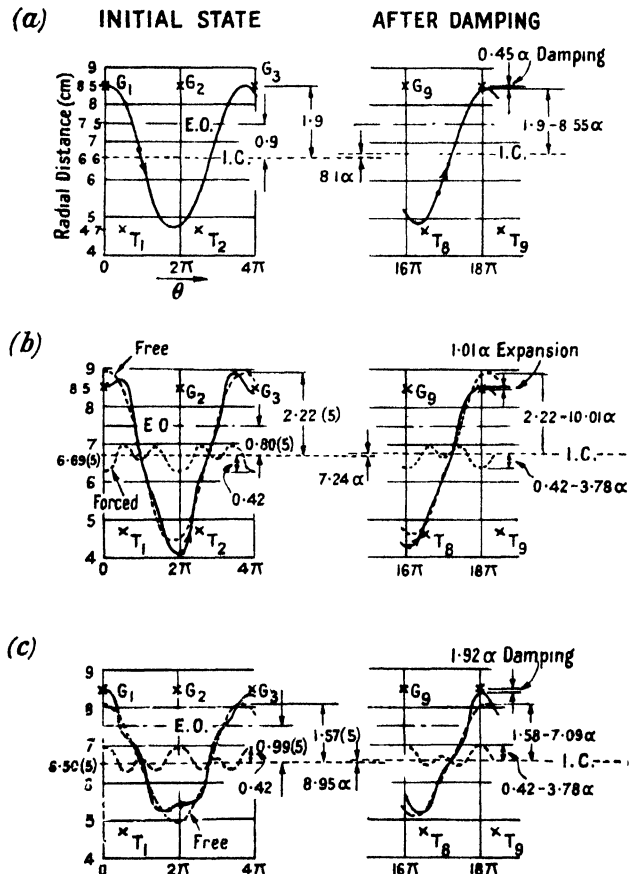


Figure 5. Effect of forced oscillation on damping.
(a) No forced oscillation; (b) harmful forced oscillation;
(c) helpful forced oscillation.

to hit the gun. The free oscillation damping is therefore 8.55α cm. From this must be subtracted the expansion of the instantaneous circle, 0.9α cm. or 8.1α cm. (The orbit expands because the equilibrium orbit is at a greater radius than the instantaneous circle.) The net damping is therefore $+0.45\alpha$ cm. These damping effects are shown in the right-hand portion of figure 5(a). The relative dampings are drawn to an approximately correct scale, but the absolute magnitude is not to scale.

The combination of a free and forced oscillation is shown in figure 5(b). The free oscillation is large and of amplitude $1.9 + \frac{1}{2}(0.65) = 2.22_5$ cm., about an instantaneous circle of radius 6.69_5 cm.; its damping after nine revolutions is therefore 10.01α cm. From this must be subtracted an expansion in the instantaneous circle of $9 \times 0.80_5\alpha = 7.24\alpha$ cm. Further, the damping of the forced oscillation, $9\alpha x_G$ or 3.78α cm., must also be subtracted. This gives a net damping of -1.01α cm. Thus the presence of the forced oscillation spoils the chance of injection.

The above effect is, however, dependent on the phase of the forced oscillation, and the forced oscillation shown in figure 5(c) may be shown to have the opposite effect. The free oscillation is now reduced to an amplitude of $1.9 - \frac{1}{2}(0.65) = 1.57_5$ cm. about an instantaneous circle of radius 6.50_5 cm. The free oscillation is damped by 7.09α cm., from which must be subtracted the expansion of the instantaneous circle 8.95α cm.; the damping of the forced oscillation, 3.78α cm., must be added in this case. The net damping is therefore 1.92α cm., which increases the chance of injection over that with no forced oscillation present.

It may be observed that the damping increase $(1.92\alpha - 0.45\alpha)$ cm., which equals 1.47α cm., is equal to the decrease $(1.01\alpha + 0.45\alpha)$ cm., to the accuracy to which the example has been worked. This change is in fact equal to the damping corresponding to one quarter the distance $x_T - x_G$ in the forced oscillation, i.e. to $9(0.65\alpha/4)$ cm., a result which may be shown to be correct in general.

From this example it is possible to generalize and say that injection may be helped or hindered by different magnitudes and angular distributions of eddy-current fields. In addition to these damping effects the betatron will clearly not work if the forced oscillation is such that it cannot be contained within the space between the gun and target, or between the donut walls, irrespective of damping.

It is a little difficult to generalize beyond this. Firstly, in the example the special case has been considered of an instantaneous circle as far from the gun as is permissible. Other possible injection conditions have been neglected, and therefore no consideration taken of the time interval over which injection is possible. Secondly, a very specific geometrical arrangement has been studied, whereas many betatrons use the gun as its own target, or have the donut walls very close to the gun or the target. It is certainly possible to observe, however, that the three types of damping shown in figure 4 are all of the same order, and that the changes introduced by various mixtures of free and forced oscillation are relatively small. Thus in the example, even if the condition of equation (14) were observed, the maximum damping change is $4.5 D_2\alpha$, whereas for injection as in figure 4(a) the damping is $9A_1\alpha$. Thus for a damping change equal to 10% of the damping of figure 4(a), we have the condition $D_2 = \frac{1}{5}A_1$, and this is therefore suggested as an upper limit to the eddy currents permissible, i.e. the overall amplitude of the forced oscillation should not exceed one-fifth of the radial separation of the gun and target.

In concluding this section it should be re-emphasized that the above considerations are not intended to indicate how normal injection into a betatron is accomplished. The assumptions of a gun which is almost a point and of the significance of damping are the recognized doubtful assumptions made by Kerst and Serber (1947). The treatment given here, while not attempting to elucidate the problem of why most betatrons work, may perhaps indicate why some do not.

§ 4. EJECTION OF ELECTRON BEAM BY A STEP FIELD

As a further example, consider the forced oscillation caused by the "step field" shown in figure 6; such a step field may be used for ejecting the electrons from their normal orbit. The disturbing field, H_D , may be analysed into Fourier components thus (note that $m=0$ does not give the $\frac{1}{4}\theta_1$ term):

$$H_D = (2\Delta H_0/\pi) [\frac{1}{4}\theta_1 - \sin \frac{1}{2}\theta_1 \cos \theta + \frac{1}{2} \sin \theta_1 \cos 2\theta \dots + (-1)^m m^{-1} \sin \frac{1}{2}m\theta_1 \cos m\theta]. \quad \dots\dots (18)$$

If it is assumed that the addition of the disturbing field does not change the value of n , then the forced oscillation is

$$x = [r_0 \Delta H_0 \theta_1] / [2\pi H_0 (n-1)] + r_0 \sum_{m=1}^{\infty} \epsilon_m \cos m\theta, \quad \dots\dots (19)$$

where
$$\epsilon_m = \frac{2}{\pi} \frac{\Delta H_0}{H_0} \left[\frac{(-1)^m \sin \frac{1}{2}m\theta_1}{m(m^2+n-1)} \right]. \quad \dots\dots (20)$$

This series converges rapidly if $\theta_1 \simeq \pi$, but less rapidly if $\theta_1 \simeq 0$ or 2π .

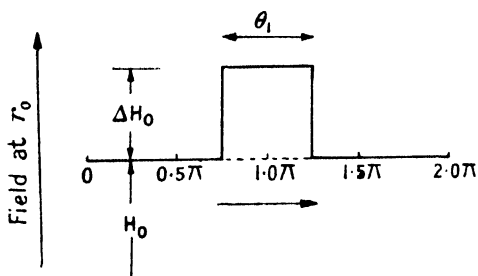


Figure 6. Step field.

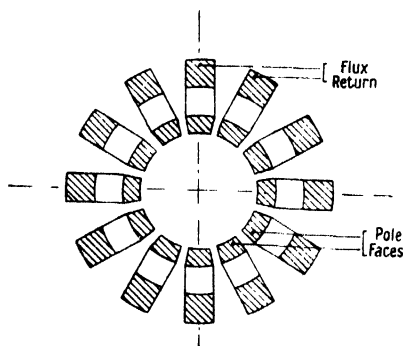


Figure 7. Circularly symmetrical magnet.

For the important case when $\theta_1 = \pi$, ϵ_2 , ϵ_4 , and even coefficients, vanish, leaving as the only significant terms

$$\epsilon_0 = \frac{2}{\pi} \frac{\Delta H_0}{H_0} \left[\frac{1}{n-1} \frac{\pi}{4} \right], \quad \epsilon_1 = \frac{2}{\pi} \frac{\Delta H_0}{H_0} \left[-\frac{1}{n} \right]. \quad \dots\dots (21)$$

Clark *et al.* (1946) considered the special case of $\epsilon_0 = 0$, and equations (19) and (21) then give their result, namely

$$\frac{x}{r_0} = -\frac{2}{\pi n} \frac{\Delta H_0}{H_0} \cos \theta. \quad \dots\dots (22)$$

For general considerations, however, such as finding the optimum length of a step field for ejecting the electrons, the slow or non-convergence of equation (19) makes it better to ignore the approach via the Fourier analysis and to treat the problem directly, as was done by Goward and Dain (1947).

§ 5. EFFECT OF PERIODIC DISCONTINUITIES IN A CIRCULARLY SYMMETRICAL MAGNET

As a final example, the effect of a large number of similar discontinuities round the circumference of a magnet may be considered. This is of some interest in the design of large synchrotrons, where it is mechanically desirable to leave gaps in the magnet poles between individual segments of the magnet. Figure 7 shows a section through such a magnet, with 12 segments. The disturbing field may then take the form shown diagrammatically in figure 8.

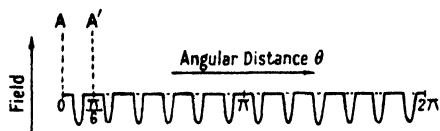


Figure 8. Possible disturbing field with circularly symmetrical magnet.

If w is the number of segments, i.e. the number of disturbances per revolution, equations (8) and (9) become

$$H_z = H_0(r/r_0)^{-n} \left[1 + \sum_{m=0}^{\infty} \delta_m \sin(mw\theta + \alpha_m) \right], \quad \dots\dots(23)$$

$$x = \left[r_0 \sum_{m=0}^{\infty} \epsilon_m \sin(mw\theta + \alpha_m) \right] + [A \sin(1-n)^{\frac{1}{2}}\theta + B \cos(1-n)^{\frac{1}{2}}\theta], \quad \dots\dots(24)$$

where $\epsilon_m = \delta_m / (m^2 w^2 + n - 1).$ \dots\dots(25)

Because of the $m^2 w^2$ term in equation (25), the forced oscillation is, in general, very much smaller for this kind of disturbance than for the same disturbance spread out so that AA' in figure 8 occupies 2π . The $m=0$ term is an exception, however, being independent of w ; this has a simple physical explanation, since it indicates a shift of the equilibrium orbit which is dependent only on the mean value of the disturbing field, taken over one revolution.

§ 6. CONCLUSION AND ACKNOWLEDGMENT

The examples which have been studied should be of interest to synchrotron and betatron designers, particularly in indicating the tolerances to which field homogeneity must be maintained during injection. The theory given may also suggest new methods of ejecting the electron beam; an instance of this will be described in the appendix.

This work was carried out in the Atomic Energy Research Establishment Group at the Telecommunications Research Establishment, Malvern, and acknowledgment is made to the Director for permission to publish this paper.

APPENDIX

EFFECT OF RADIAL DISTURBING FIELDS

In the main text we have studied the effects of azimuthal asymmetries, formed by the addition of axial fields to the main field. The problem of adding radial fields to give azimuthal asymmetry may be considered in a very similar way.

The undisturbed radial field of a betatron or synchrotron may be obtained by applying the condition that curl H is zero, i.e. that

$$\partial H_z / \partial r = \partial H_r / \partial z. \quad \dots\dots(26)$$

This gives, for the radial magnetic field,

$$H_r = -nzH_0/r_0, \quad \dots\dots(27)$$

to a good approximation. Consider, now, the addition of a radial field, varying sinusoidally with azimuth, so that the field becomes

$$H_r = -nzH_0/r_0 + H_0\beta_m \sin(m\theta + \alpha_m). \quad \dots\dots(28)$$

The equation of axial motion, for an electron of velocity v_0 , is

$$d^2z/dt^2 = (e/mc)v_0H_r, \quad \dots\dots(29)$$

which reduces, using equations (2) and (28), to

$$d^2z/d\theta^2 = -nz + r_0\beta_m \sin(m\theta + \alpha_m). \quad \dots\dots(30)$$

The general solution of this equation is

$$z = [-r_0\gamma_m \sin(m\theta + \alpha_m)] + [C \sin(n\theta) + D \cos(n\theta)], \quad \dots\dots(31)$$

where

$$\gamma_m = \beta_1/(m^2 - n). \quad \dots\dots(32)$$

Arbitrary disturbing fields may be synthesized by expressing them as Fourier series, as in the main text.

It may be seen, from equations (9) and (31), that an azimuthally constant disturbing field, for which $m=0$ and $\alpha_m=\pi/2$, gives a displacement $(1/n)-1$ times as great as would the same field applied axially: this is in agreement with the well known axial instability observed as n approaches zero.

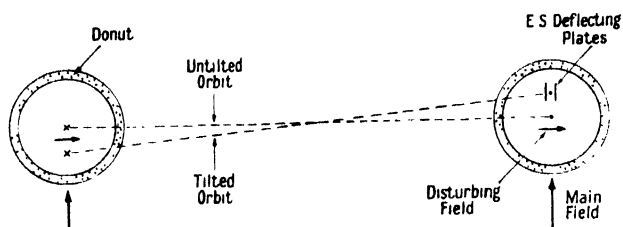


Figure 9. Beam ejection by radial disturbing field.

The effect of a disturbing field varying as $\sin \theta$, i.e. $m=1$, is of particular interest and importance. The orbit is caused to tilt, its maximum displacement from $z=0$ being $r_0\beta_1/(1-n)$. As n approaches 1, the displacement becomes resonant and very large. This large displacement suggests a method of extracting the beam in a betatron or synchrotron, since the beam could be made to pass into electrostatic deflecting plates. This is shown diagrammatically in figure 9.

REFERENCES

- ADAMS, G. D., KERST, D. W., and SCAG, D. T., 1947, *Rev. Sci. Instrum.*, **18**, 799.
 BOHM, D., and FOLDY, L., 1946, *Phys. Rev.*, **70**, 249.
 CLARK, J. S., GETTING, I. A., and THOMAS, J. E., JR., 1946, *Phys. Rev.*, **70**, 562.
 GOWARD, F. K., and DAIN, J., 1947, *Nature, Lond.*, **159**, 636.
 KERST, D. W., 1941, *Phys. Rev.*, **60**, 47.
 KERST, D. W., and KOCH, H. W., 1947, *Rev. Sci. Instrum.*, **18**, 681.
 KERST, D. W., and SERBER, R., 1941, *Phys. Rev.*, **60**, 53.

LETTERS TO THE EDITOR

The Measurement of Dielectric Constant in a Waveguide

The dielectric constant of ethyl alcohol has been measured at a wavelength of 12.60 cm. in a waveguide. The standing-wave method was used, the essentials of which have already been described by Williams and Bolton (1945). The wavelength is close to the middle of the dipole dispersion region and the absorption is consequently large. The energy is completely absorbed in the liquid when a few centimetres deep, and the impedance presented at the liquid surface is the characteristic impedance of the guide filled with dielectric.

The liquid was supported on a thin slab of ebonite in a vertical guide, 7×3.5 cm. in cross-section, and an H_{10} wave was used. The minimum depth of ethyl alcohol needed was about 4 cm.: for depths greater than this the impedance measured was constant. The impedance is complex and, using the symbols defined by Williams and Bolton, is given by $Z_0' = i\omega\mu/P_t'$ where $P_t' = \alpha_t' + i\beta_t'$ is the propagation constant. If χ^2 is the constant characterizing the mode of oscillation in the hollow tube, $\chi^2 = P_t'^2 + \omega^2\mu(\epsilon' - i\epsilon'')$, and these three equations are sufficient to determine the real and imaginary parts (ϵ' , ϵ'') of the dielectric constant.

The standing wave was observed by a crystal detector moving in a slot above the liquid. The crystal was calibrated by assuming its characteristic to be given by $i = A|E|^2/k$, where i is the output current and E the strength of the transverse electric field. The value of k was found by a subsidiary experiment using an empty tube closed by a metal plate. The standing wave was then represented by $|E|^2 = B \sin^2 2\pi x/\lambda_t$, where x is the distance from a minimum, and the slope of the graph of $\log i$ against $\log (\sin 2\pi x/\lambda_t)$, which was accurately a straight line, gave k equal to $0.960 \pm 0.5\%$.

The results of three independent samples of freshly dried and distilled ethyl alcohol gave $\epsilon' = 5.5$, $\epsilon'' = 7.8$. The accuracy was to within about 1%. The results are plotted in the figure together with other experimental points referred to by Mizushima (1927), Akerlof (1932), Malsch (1932), Báz (1939) and Slevogt (1939). The semicircle is the locus of Debye's equation and the points suggest that the equation applies to ethyl alcohol. Using the parametric form of the equation the present experimental results give a value of 31 cm. for the relaxation wavelength (Sprungwellenlänge).

The University, Nottingham.

H. C. BOLTON.

12 June 1948.

AKERLOF, G., 1932, *J. Amer. Chem. Soc.*, **54**, 4125.

BÁZ, G., 1939, *Phys. Z.*, **40**, 394.

MALSCH, J., 1932, *Ann. Phys., Lpz.*, **12**, 865.

MIZUSHIMA, S., 1927, *Phys. Z.*, **28**, 418.

SLEVOGT, K. E., 1939, *Ann. Phys., Lpz.*, **36**, 141.

WILLIAMS, G., and BOLTON, H.C., 1945, *Phil. Mag.*, **36**, 862.

REVIEWS OF BOOKS

Electrons in Gases, by Sir JOHN TOWNSEND, F.R.S. Pp. 166. (London: Hutchinson's Scientific and Technical Publications, 1948.) 25s.

Sir John Townsend, the author of this book, by the number and fundamental character of his contributions, both in theory and experiment, occupies an unrivalled position as an authority in the field of electrical discharges and of electronic motion in gases.

Moreover, his early investigations at the turn of the century on the atomic nature of electricity rank with the greatest of that remarkable time. It will be recalled, for instance, that in 1897 he made the first direct attempt to determine the atomic charge by the method of

falling drops, which, to quote Millikan, "was of much novelty and no little ingenuity. It is also of great interest because it contains all the essential elements of some subsequent determinations." In 1899 and 1900 he showed from the ratio of the mobilities of ions to their diffusion coefficients (which he measured) that the product is the same for gaseous ions as for a monovalent ion in an electrolyte. Thus the charge e is the same on each. These fundamental investigations were followed in 1900 by his discovery that molecules of a gas were ionized in collisions with electrons possessing energies of only 20 electron volts, and by means of his well-known parallel plate apparatus he measured the number of molecules α ionized on the average by an electron in drifting one centimetre in the direction of the uniform electric force. These early experiments on ionization possess a double interest; first, they showed that electrons liberated by ultra-violet light from a metal plate and those liberated from molecules by ionization possessed the same ionizing coefficient α and were presumably identical—a fact of considerable significance at that time; secondly, they form the starting point of the series of investigations on the mechanism of discharges in gases to which we owe most of our understanding of the subject. When n_0 photo-electrons are liberated from a plane cathode, the number reaching the anode after moving in a uniform field between the electrodes is given by Townsend's formula

$$n = n_0 \frac{(\alpha - \beta) \exp(\alpha - \beta)x}{\alpha - \beta \exp(\alpha - \beta)x},$$

in which α has the meaning assigned to it above, and β is the number of molecules ionized by a positive ion in drifting one centimetre in the direction of the electric force which is supposed to be maintained constant as the separation x of the electrodes is varied. Thus in 1903 Townsend was led to propose a theory of sparking in uniform fields according to which a spark first occurs at a separation x such that the denominator in the expression vanishes. The current then remains finite although n_0 is zero. Since he had shown, both by theory and measurement, that the ionizing coefficients α and β in each gas satisfy functional equations $\alpha/p = f(Z/p)$ and $\beta/p = F(Z/p)$, p being the gas pressure and Z the electric force, it was possible successfully to predict the sparking potential for a given separation x of the electrodes from the experimental determinations of these functional relationships. In the following years up to the outbreak of the first World War, the theory was extended to discharges in non-uniform fields and to the conditions that supervene when the applied field is modified by the presence of space charge. This work is lucidly described in his celebrated treatise *Electricity in Gases*, which, in many respects, is still the best account of the principles of the subject.

The appearance of a recent book, *Electrons in Gases*, by Sir John Townsend, describing his further investigations in this field carried out between the wars, is, therefore, a matter of considerable interest to workers in this and related subjects. Although the book forms a useful appendix to *Electricity in Gases*, it is in fact a self-contained work because the author has included all the theory required for an understanding of the whole book. The book is principally concerned with three topics: the study of the behaviour of slow electrons in gases when ionization is unimportant, ionization by collision, and the theory of the spark discharge. After a preliminary chapter which gives the theory of electron diffusion and drift in an electric field, the author proceeds to describe the principle of his diffusion apparatus, which permits the energy factor k —the ratio (agitational energy of electrons)/(agitational energy of the gas molecules)—and the drift velocity W of the electrons to be determined as a function of the ratio (Z/p) of the electric force Z to the pressure p . From k and W , using theoretical formulae, the following parameters of the electronic motion are obtained: the mean free path L at unit pressure (1 mm. Hg), the r.m.s. velocity of agitation V and the mean proportion λ of its energy lost by an electron in a collision with a gas molecule. The dependence of k , W , V and λ upon Z/p in a variety of gases is exhibited in tables and curves.

It was found that L (which is inversely proportional to the collisional cross-section of the molecules) depends upon the energy of the electron, the variation being especially remarkable in the heavier monatomic gases. There are interesting critical paragraphs in which the results of diffusion experiments on L are compared with those of other workers using the methods of Lenard and Ramsauer.

The major part of the remainder of the book is concerned with studies of ionization in gases by collision and other actions, and of theories of discharges. It has long been known

(*Electricity in Gases*) that the increase of the current with distance in a uniform field is represented with equal accuracy both by the formula given above in terms of the coefficients α and β and by an alternative formula in terms of α and γ , where the positive ions are supposed to act by liberating γ electrons per ion at the cathode, and also by a formula in which α , β and γ all occur. It is a remarkable fact that all three formulae are analytically equivalent, so that it is not possible to decide the precise action of the positive ions in contributing to the growth of the current and in electrical discharges, from experiments with uniform fields. When, however, the whole range of phenomena is considered, in particular the properties of brush discharges from wires and points, the author shows that the fashionable view that the action represented by the coefficient β is absent is untenable, but that it is necessary to suppose that positive ions do in fact ionize gas molecules in discharges—that is, the β -effect is always operative. Near the minimum sparking potential the γ -action is also important, but becomes relatively unimportant at the higher pressures. He also describes an experiment which demonstrates the β -action directly.

Other theories of sparking are discussed in full and the hypotheses on which they are based are shown to be untenable. Although a theory of sparking at present in vogue is not discussed, it is rejected by implication, since it is shown that photo-ionization of the gas is not an important factor in contributing to electrical breakdown. The book contains interesting and acute discussions of many other matters, including the electrical properties of monatomic gases, experiments on critical potentials, and the distribution of the agitational velocities of the electrons.

This is a book which should be read by all who are interested in the fundamental principles of electrical discharges and of electronic motion in gases.

L. G. H. HUXLEY.

Royal Society Mathematical Tables Committee

Mathematical Tables for Science and Industry

It is considered that important investigations in the physical, chemical and engineering sciences, in mathematics and in industry, may be held up by the absence of relevant mathematical tables. Several important tables of a fundamental nature have in the past been produced by the British Association through its Mathematical Tables Committee. The work in this field has now been transferred to the Royal Society on the invitation of the British Association, and a special Royal Society Mathematical Tables Committee has been established to continue and extend the earlier work. This Committee would be glad to receive suggestions from industrial research groups, from individual investigators and from Service departments relating to existing needs in their special fields. Requests for assistance may be met either by advice as to the most economical way of producing the desired tables, or by an undertaking to produce the tables under the auspices of the Royal Society, or, in exceptional cases, by a grant towards the cost of tabulation or to make possible the publication of important tables which may exist only in manuscript. In the first instance inquiries should be addressed to the Assistant Secretary, The Royal Society, Burlington House, London W.1.

CORRIGENDA

“Angular Distribution in Internal Pair Creation”, by G. K. HORTON (*Proc. Phys. Soc.*, 1948, **60**, 457).

In figures 1, 2, 3, ordinates *should read* $P \times 10^5$, *not* $P \times 10^{-5}$.

In figure 1, right-hand ordinate scale should be shifted down one unit.

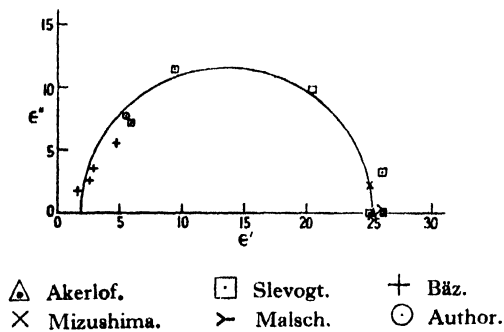
In figure 4, caption *should read*: “2, 4, Magnetic dipole or electric quadrupole. 1, 3, Electric dipole.”

Page 460, line 11, *for larger read smaller*.

CORRIGENDUM

“The Measurement of Dielectric Constant in a Waveguide”, by H. C. BOLTON
(*Proc. Phys. Soc.*, 1948, **61**, 294).

The following figure should have been inserted in the above Letter to the Editor.



THE PROCEEDINGS OF THE PHYSICAL SOCIETY

VOL. 61, PART 4

1 October 1948

No. 346

On Schwinger's Theory of Nuclear Forces

By W. H. RAMSEY

Turner and Newall Research Fellow, University of Manchester

MS. received 7 May 1948

ABSTRACT. It is shown that Schwinger's generalization of the Møller-Rosenfeld theory cannot consistently be used to describe nuclear forces. It does not lead to the correct quadrupole moment of the deuteron, and it offers no hope of a detailed agreement with proton-proton scattering.

§ 1. INTRODUCTION

THE symmetrical version of the meson theory of nuclear forces proposed by Møller and Rosenfeld (1940) has been used with some success to calculate the binding energies of the very light nuclei (Frohlich, Huang and Sneddon 1947) and the features of neutron-proton scattering (Ramsey 1947). The meson mass found by the former authors from the energy levels of the hydrogen isotopes is $210\text{--}220 m_e$, where m_e is the mass of the electron; this mass is consistent with the results of neutron-proton scattering and it is in fair agreement with the mass of the commonest cosmic-ray meson. On the other hand, Ramsey (1948) has shown that the empirical data on proton-proton scattering set a lower limit of $275 m_e$ to the mass of the meson which must be used in this theory. The failure of the theory can be traced to the radial dependence of the interaction. Schwinger (1942) has proposed a generalization of the Møller-Rosenfeld theory which retains the essential characteristics of the simple theory and, at the same time, the interaction has a different radial dependence. The purpose of this note is to examine the possibility of using Schwinger's theory to give a consistent description of nuclear forces.

In Schwinger's theory the static interaction between two nucleons is

$$\mathbf{V}(\mathbf{r}) = \mathbf{V}_v(\mathbf{r}) + \mathbf{V}_{ps}(\mathbf{r}), \quad \dots\dots(1)$$

where $\mathbf{V}_v(\mathbf{r})$ is the interaction in the symmetrical vector theory

$$\mathbf{V}_v(\mathbf{r}) = (\boldsymbol{\tau}_1, \boldsymbol{\tau}_2) \{ g_1^2 + \frac{2}{3} g_2^2 (\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2) - (3\lambda^2 r^2)^{-1} g_2^2 \mathbf{S}_{12} (3 + 3\lambda r + \lambda^2 r^2) \} r^{-1} \exp(-\lambda r), \quad \dots\dots(1a)$$

and where $\mathbf{V}_{ps}(\mathbf{r})$ is the interaction in the symmetrical pseudoscalar theory

$$\mathbf{V}_{ps}(\mathbf{r}) = (\boldsymbol{\tau}_1, \boldsymbol{\tau}_2) \{ \frac{1}{3} f_2^2 (\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2) + (3\kappa^2 r^2)^{-1} f_2^2 \mathbf{S}_{12} (3 + 3\kappa r + \kappa^2 r^2) \} r^{-1} \exp(-\kappa r). \quad \dots\dots(1b)$$

In these equations $\mathbf{S}_{12}$ denotes the dipole-dipole operator. In order to suppress the r^{-3} singularity in (1) Schwinger imposes the subsidiary condition

$$(g_2/\lambda)^2 = (f_2/\kappa)^2. \quad \dots\dots(2)$$

The Møller-Rosenfeld theory appears as the special case in which the masses of the vector and pseudoscalar mesons are equal (that is, $\lambda = \kappa$).

§ 2. PROTON-PROTON SCATTERING

The scattering of protons by protons provides a means of examining the radial dependence of the interaction in the ^1S state; in this configuration the complication of non-central forces does not arise. At low energies the electrostatic repulsion prevents the close approach of the protons so that the anomalous scattering is controlled by the nuclear interaction at large distances. On the other hand, the interaction at small distances is very important in proton-proton scattering at high energies and in neutron-proton scattering. Using the Møller-Rosenfeld potential,

$$V(r) = -Gr^{-1} \exp(-\mu r), \quad \dots\dots(3)$$

Ramsey (1948) found that the empirical data on proton-proton scattering require a meson mass of at least $275 m_e$. There is, however, no detailed agreement with experiment for any mass of the meson. The data for energies below 2 MeV. suggest a mass of $325 m_e$, whilst the results at higher energies favour a mass of $270 m_e$. Thus, when the constants of (3) are adjusted to give a good description of the nuclear interaction at a distance r_0 , the results of proton-proton scattering suggest that the value of μ increases with r_0 .

In Schwinger's theory the interaction in the ^1S state is

$$V(r) = -(f_2^2 + 2g_2^2 - g_1^2)\{(1-\alpha)\exp(-\lambda r) + \alpha\exp(-\kappa r)\}r^{-1}, \quad \dots\dots(4)$$

where

$$\alpha = f_2^2/(f_2^2 + 2g_2^2 - g_1^2).$$

If the constants of (3) are chosen to give a good agreement with (4) in the neighbourhood of r_0 , then

$$\mu = \{(1-\alpha)\lambda \exp(-\lambda r_0) + \alpha\kappa \exp(-\kappa r_0)\} / \{(1-\alpha)\exp(-\lambda r_0) + \alpha\exp(-\kappa r_0)\} \quad \dots\dots(5)$$

and

$$d\mu/dr_0 = -\alpha(1-\alpha)(\lambda-\kappa)^2\{(1-\alpha)\exp(-\lambda r_0) + \alpha\exp(-\kappa r_0)\}^{-2} \exp(-(\lambda+\kappa)r_0).$$

It will be noted that the sign of $d\mu/dr_0$ depends only on α ; hence μ is a monotonic function of r_0 . The cases $\alpha=0$ and $\alpha=1$ are trivial as (3) and (4) become identical. The fact that (4) must be attractive rules out negative values of α .

§ 3. CASE I: $0 < \alpha < 1$

The range $0 < \alpha < 1$ corresponds to the case $2g_2^2 > g_1^2$. When this condition is fulfilled, μ as given by (5) is a continually decreasing function of r_0 . Under these circumstances the interaction (4) may be expected to be less satisfactory than (3) when applied to proton-proton scattering.

Detailed calculations of the properties of the deuteron have been made by Jauch and Hu (1944) on the basis of Schwinger's theory with the subsidiary conditions

$$g_1^2 = 0, \quad \dots\dots(6a)$$

$$\hbar\kappa/c = 177 m_e. \quad \dots\dots(6b)$$

The charge f_2 and the mass of the vector meson were determined from the ^3S and ^1S energy levels; the latter quantity was found to be 60% greater than (6b). The theory can then account for only one-third of the observed quadrupole moment; the ratio λ/κ would have to be increased to about 10 to give the empirical value. If g_1 is assumed different from zero, the discrepancy becomes even worse.

As $\mu = \kappa$ for large values of r_0 , a pseudoscalar meson mass smaller than (6b) cannot be contemplated; it should, in fact, be increased to the neighbourhood of $300 m_e$. This change will exaggerate the disagreement found by Jauch and Hu. It would therefore appear that Schwinger's theory with $0 < \alpha < 1$ is incapable of accounting for either proton-proton scattering or the properties of the deuteron.

§ 4. CASE II: $\alpha > 1$

The case $\alpha > 1$ corresponds to the following restrictions on the charge constants:

$$f_2^2 > g_1^2 - 2g_2^2 > 0. \quad \dots\dots(7)$$

Under these circumstances μ is a continually increasing function of r_0 , and a detailed agreement with the results of proton-proton scattering may be obtainable. The possibility that (4) becomes repulsive at large distances must be excluded as this would be at variance with the results of experiments on the scattering of very slow protons by protons; this imposes the further restriction that the mass of the vector meson must be the larger:

$$\lambda > \kappa. \quad \dots\dots(8)$$

As the maximum value of μ is equal to κ , the mass of the pseudoscalar meson must be at least $275 m_e$.

The magnitude of the interaction constant G obtained by comparing (3) and (4) increases monotonically with r_0 to the limiting value f_2^2 . In order to account for the values of G determined empirically, f_2^2 must be at least $0.25 h c$. Relations (2) and (7) then set a lower limit to g_1^2 :

$$g_1^2 > 0.5(\lambda/\kappa)^2 h c.$$

These abnormally large charge constants, together with condition (8), prevent the theory from leading to the correct binding energy of the deuteron. The magnitude of g_1^2 also renders the use of the static interaction (1a) unreliable, even as a first approximation.

§ 5. CONCLUSION

Schwinger's generalization of the Møller-Rosenfeld theory cannot consistently be used to describe nuclear interactions. It does not lead to the correct quadrupole moment of the deuteron, and it offers no hope of a detailed agreement with proton-proton scattering.

ACKNOWLEDGMENT

The author would like to express his appreciation to Professor L. Rosenfeld for the interest he has shown in the work.

REFERENCES

- FRÖHLICH, H., KUN HUANG, and SNEDDON, I. N., 1947, *Proc. Roy. Soc. A*, **191**, 61.
 JAUCH, J. M., and NING HU, 1944, *Phys. Rev.*, **65**, 289.
 MØLLER, C., and ROSENFELD, L., 1940, *Kgl. Danske Vid. Selsk. mat.-fys. Meddl.*, **17**, No. 8.
 RAMSEY, W. H., 1947, *Proc. Roy. Soc.*, **191**, 195; 1948, *Ibid.*, in press.
 SCHWINGER, J., 1942, *Phys. Rev.*, **61**, 387.

The Delayed Coincidence Method in the Study of Radioactivity, with Application to Isomerism in ^{181}Ta

By D. E. BUNYAN, A. LUNDBY, A. H. WARD AND D. WALKER

Physics Department, University of Birmingham

*Communicated by M. L. Oliphant; MS. received 19 February 1948,
and in amended form 12 May 1948.*

ABSTRACT. A short-lived nuclear state occurring during a series of radioactive transitions can be studied by the observation of delayed coincidences between the radiations preceding and following the state in question. The method is limited at present to transition periods lying between about 10^{-7} and 10^{-2} seconds. Two distinct experimental procedures are possible, one yielding the differential and the other the integrated decay curve of the short-lived state.

The method has been applied to the decay of the previously reported 22-microsecond metastable state in ^{181}Ta , which is formed in the decay of the β -ray emitter ^{181}Hf . It is shown that the isomeric transition takes place in two steps of 0.2 mev. and 0.5 mev. in instantaneous succession. The half-life of the metastable level has been redetermined as $(20.1 \pm 0.7) \times 10^{-6}$ seconds. From a comparison of the observed and calculated internal conversion coefficients, it seems probable that the 0.2 mev. and 0.5 mev. transitions are of an electric octopole and electric quadrupole character respectively. This would mean that since the spin of the ground state of ^{181}Ta is $7/2$, the 0.5 mev. state must also have a spin of $7/2$ and the same parity as the ground state, while the metastable state must have a spin of $1/2$ and opposite parity.

§ 1. INTRODUCTION

WHERE an atomic nucleus undergoes a series of radioactive transitions following rapidly upon one another, coincidence observations on the radiations emitted, using electrical counting techniques, can yield information about the time relationships of the processes involved. An obvious practical restriction is that the initial transition must be of long period.

If one transition follows another within less than about 10^{-8} seconds, it is said to give rise to "instantaneous" coincidences, since experimental coincidence arrangements with the best possible resolution attainable at present* are unable to indicate any departure from simultaneity. On the other hand, if the second transition follows the first after a mean time interval somewhat greater than 10^{-8} seconds, it is possible to detect that the two radiations emitted are not wholly in coincidence. Where the mean time interval lies between about 10^{-7} and 10^{-2} seconds, it can be measured by delaying or lengthening the electrical pulse recording the first transition so as to bring it into coincidence with that for the second.

The procedure where the first pulse is delayed yields the differential decay curve of the intermediate short-lived state (DeBenedetti and McGowan 1946), while that where the first pulse is lengthened yields the integrated decay

* using Geiger counters.

curve (Dunworth 1939, Rotblat 1941). The elements of a delayed coincidence recorder are illustrated in figure 1.

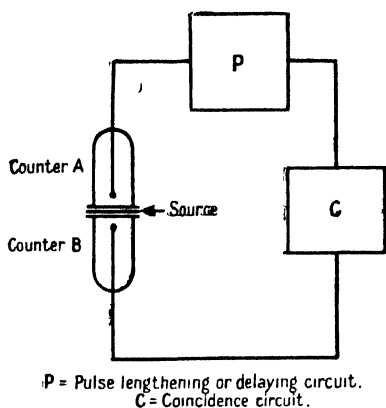


Figure 1. Elements of delayed coincidence recorder.

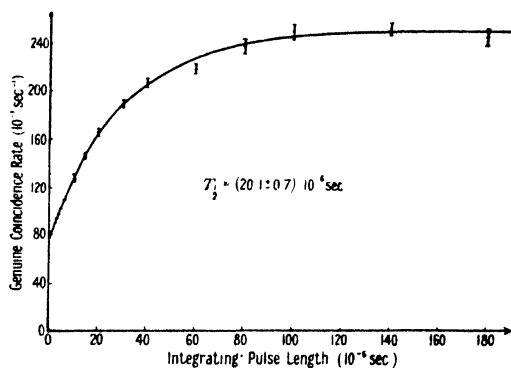


Figure 2. The integrated decay curve of the metastable (upper) level of ^{181}Ta .

§ 2. COMPARISON OF THE DIFFERENTIAL AND INTEGRAL PROCEDURES

(i) Ratio of genuine to random coincidence rates

Measurements with any coincidence system are limited by the presence of random coincidences. If counters A and B (figure 1) are irradiated by a source placed between them to give individual counting rates of N_A and N_B respectively and a genuine coincidence rate C_g , there will be in addition a background rate of random coincidences given by $C_r = (N_A - C_g)(N_B - C_g)(\tau_A + \tau_B)$, where τ_A and τ_B are the lengths of the pulses reaching the coincidence circuit from channel A and channel B respectively. Since in radioactive experiments C_g is in general small compared with N_A or N_B , we can write, without appreciable error,

$$C_r = N_A N_B (\tau_A + \tau_B). \quad \dots\dots (1)$$

Experimentally one finds C_g by determining the total coincidence rate, and subtracting C_r , which is computed from a knowledge of N_A , N_B , and $(\tau_A + \tau_B)$ using equation (1). It is clearly essential to have C_g/C_r sufficiently large.

In the integral procedure for delayed coincidences, τ_A is increased progressively while plotting the decay curve, and it can be seen that the ratio C_g/C_r decreases for increasing τ_A since C_r increases linearly with τ_A , whereas C_g increases typically as $\{1 - \exp(-\lambda\tau_A)\}$, where λ is the decay constant of the short-lived state. (Any superimposed instantaneous coincidences can be suppressed by the addition of an anti-coincidence circuit.) Similarly in the differential procedure C_g/C_r decreases for increasing delay t_d in the A channel, since although C_r remains constant, C_g decreases typically as $\exp(-\lambda t_d)$. This decrease of C_g/C_r sets an upper limit of about 10^{-2} seconds to half-lives which can be measured by the integral procedure, as has been pointed out by Rotblat (1941). The upper limit to half-lives measurable with the differential procedure is lower since it may be shown that C_g/C_r decreases more rapidly with increasing delays than is the case for the integral procedure.

(ii) *Time-scale calibration*

Time calibration in the integral procedure can be carried out simply with random coincidences using equation (1). For this purpose the two counters are irradiated by separate sources to give independent counting rates. Time calibration in the differential procedure requires accurate measurement of the time interval between pulses in the two input channels and is more complicated.

(iii) *Lower limit of measurable half-lives*

With a Geiger counter there is a variable time-lag between the entry of a particle into the counter and the detection of the output voltage pulse. Either delayed coincidence procedure fails when the half-life to be measured is rather smaller than the time-lag fluctuations exhibited by the counters. This sets a lower limit of about 10^{-7} seconds to half-lives which can be measured. With extreme care the method may be pushed to half-lives of a few times 10^{-8} seconds (Bittencourt and Goldhaber 1946), but any further improvement will have to await the development of satisfactory fast-response detectors for use in place of Geiger counters.

§ 3. EXPERIMENTAL DELAYED COINCIDENCE RECORDERS

Two such equipments have been constructed in this laboratory according to the principles shown in figure 1. One, the "differential recorder", employs a pulse-delay network (artificial line) in channel A giving a time delay variable from 0 to 10×10^{-6} seconds in steps of 0.1×10^{-6} seconds. The pulses reaching the coincidence circuit from either channel are fixed in length at 0.22×10^{-6} seconds. The other, the "integral recorder", employs a pulse-lengthening circuit (biased flip-flop) in channel A, giving pulses variable in length from 10^{-6} seconds upwards. The pulse reaching the coincidence circuit from channel B is in this case fixed in length at 0.7×10^{-6} seconds.

During the development of the integral recorder its performance was checked by redetermining the half-life of RaC' , counter B in this case being a proportional counter. The value obtained, $(1.47 \pm 0.07) \times 10^{-4}$ seconds, is in good agreement with previous measurements (Dunworth 1939, Rotblat 1941).

§ 4. EXPERIMENTS ON THE DECAY OF ^{181}Hf (i) *Previous results*

It has been reported that β -emission by ^{181}Hf is followed by an isomeric transition in ^{181}Ta , the half-life of the metastable state being 22 microseconds (DeBenedetti and McGowan 1946). Since, however, these authors did not account for some of the radiations present, it was decided to make a closer study of the disintegration scheme. Madansky and Wiedenbeck (1947) have reported, from intensity measurements based on absorption curves of the total radiation, that there appear to be two γ -rays in cascade.

(ii) *Preparation of the source, and experimental arrangements*

There was available for the investigation a mixture of hafnium oxide and zirconium oxide in the proportion of 7:3 by weight. This was in the form of a fine powder with a primary grain size of $<10^{-3}$ cm. A sample was irradiated

in the low energy pile at the Atomic Energy Research Establishment, Harwell, to give a specific activity of about 100 microcuries per gram, corresponding to a slow neutron absorption cross-section for ^{180}Hf of the order of 10^{-23} cm^2 . Experiments were commenced two weeks after the removal of the sample from the pile. Only one activity was detected, and this decayed over a half-life with a period of 48.5 ± 0.5 days. The half-life of ^{181}Hf has been given in the literature as 55 days, but Seren, Friedlander and Turkel (1947) give 46 ± 3 days as the result of a recent redetermination.

The source used in the coincidence experiments described below was made by depositing some of the fine powder on one side of an aluminium foil (2.3 mg/cm^2), the mean weight of the powder and adhesive material amounting to about 1 mg/cm^2 . Care was taken to eliminate grain conglomerates as much as possible. The foil was mounted between two end-window β -ray counters 1 cm. apart (figure 1), the source facing counter B (7 mg/cm^2 window). In the delayed coincidence experiments, counter A had a 20 mg/cm^2 window. Thus the limiting energy of electrons which could enter counter A was about 140 keV. and that for counter B about 75 keV.

In all the measurements described below corrections have been applied to allow for the decay of the source. In the coincidence absorption measurements, back-scattering effects were found to be small.

(iii) *Half-life of the metastable level of ^{181}Ta*

Figure 2, obtained using the integral recorder, shows the number of genuine coincidences between the particles detected in counter A and the particles detected in counter B as a function of the integrating pulse length. The observed genuine coincidence rate C_g can be represented by a curve of the form $C_g = C_{\text{inst}} + C_{\text{max}}(1 - e^{-\tau})$, where C_{inst} is the instantaneous coincidence rate and C_{max} the limiting delayed coincidence rate for increasing pulse length τ . The best curve of this form fitting the observations was found by the method of least squares, the value of τ corresponding to each observed point being assumed correct. This assumption is justified since, although in our experimental arrangement any individual setting of τ is subject to an uncertainty of the order of 2%, such setting errors are random in nature and the coincidence rate for each of the fifteen observed points is the average of not less than four independent settings of the required value of τ . The half-life of the metastable level of ^{181}Ta was found to be $(20.1 \pm 0.7) \times 10^{-6}$ seconds.

As regards the instantaneous coincidences exhibited in figure 2, it was shown, using the differential recorder, that the transitions involved followed one another within less than 10^{-7} seconds.

(iv) *Radiation absorption curves*

Absorption curve of the radiation preceding the isomeric transition. While plotting this curve, and the absorption curve of the delayed radiations (see below), we used the integral recorder with the integrating pulse length set at 30×10^{-6} seconds. The instantaneous coincidences were suppressed by the addition of an anti-coincidence circuit of 10^{-6} seconds resolving time.

The absorption curve of the radiation preceding the isomeric transition (A, figure 3) was obtained by putting aluminium foils in front of counter A.

It shows only one component, a β -ray of maximum energy about 0.45 mev. DeBenedetti and McGowan (1946) give the energy of the β -ray as 0.8 mev., which is much higher than can be justified from our results. Madansky and Wiedenbeck (1947), from an absorption curve of the total radiation, give 0.45 mev.

Absorption curve of the delayed radiations. The absorption curve of the delayed radiations (B, figure 3) was obtained by putting aluminium foils between the source and counter B. The absorption curve shows the presence of three components: (i) an electron component of maximum range 22 mg/cm² corresponding to 130 kev., (ii) an electron component of maximum range 150 mg/cm² corresponding to 420 kev., and (iii) a hard component (γ -radiation).

The delayed radiation absorption curve of DeBenedetti and McGowan (1946), as far as it goes, agrees with ours, but it has not been carried far enough

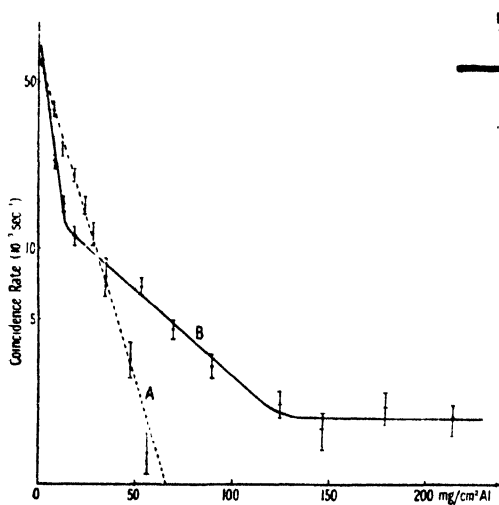


Figure 3.

A: absorption curve of radiation preceding the isomeric transition.

B: absorption curve of delayed radiation.

To obtain the total absorber thickness, the counter-window thickness must be added (20 mg/cm² for A and 7 mg/cm² for B).

to identify the end point of the harder electron component or to show the presence of the γ -ray component in the delayed radiation. They give an energy of 110 kev. for the softer electron component.

Allowing for the K binding energy of Ta (70 kev.), the electron components found by us correspond to the conversion of γ -rays of 0.2 and 0.5 mev.

Absorption curve of the radiations in instantaneous coincidence. An absorption curve of the radiations in instantaneous coincidence was taken with a resolving time of 0.44×10^{-8} seconds. Both counters had 7 mg/cm² windows in this case. This absorption curve showed the presence of the same three components as are found in the delayed radiations.

Absorption curve in lead of the γ -radiation. An extended γ -ray absorption curve taken on a separate counting set without coincidence and using lead absorbers

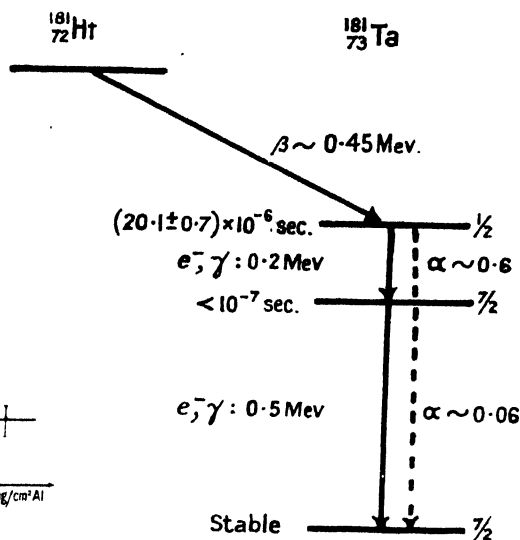


Figure 4. Decay scheme for $^{181}\text{Hf} \rightarrow ^{181}\text{Ta}$.

showed the presence of a γ -ray of 0.5 mev., which was also found in the same way by DeBenedetti and McGowan (1946). From similar measurements Madansky and Wiedenbeck (1947) give the energy as 0.52 mev., and also assign an energy of 0.30 mev. to the softer γ -ray.*

§ 5. THE DECAY SCHEME FOR $^{181}\text{Hf} \rightarrow ^{181}\text{Ta}$

Consideration of the delayed coincidence and absorption measurements leads to the following conclusions: (i) the emission of a 0.45 mev. β -particle by ^{181}Hf leads to the 20 microsecond metastable state of ^{181}Ta ; (ii) the metastable state of ^{181}Ta decays to the ground state by means of two successive transitions of 0.2 and 0.5 mev. which follow one another within 10^{-7} seconds. The second conclusion follows from the fact that the absorption curves of the delayed radiations and the radiations in instantaneous coincidence show the same components.

This leads to the proposed decay scheme shown in figure 4. No transitions from the metastable state direct to the ground state have been observed, and they cannot amount to more than a few per cent of the two-step transitions.

The internal conversion coefficient α is defined for each transition as the ratio of the number of conversion electrons to the number of conversion electrons plus γ -rays. The value of α for the 0.2 mev. transition (0.6) has been estimated from the ratio of the conversion electron and β -ray intensities obtained from a detailed absorption curve of the total radiation emitted by the Hf sample, using a counter with a 2.3 mg/cm² window. The value of α for the 0.5 mev. transition (0.06) has been further estimated from the relative intensities of the two conversion electron components in curve B, figure 3.

The multipole order of the 0.2 mev. and 0.5 mev. transitions, their sequence, and the probable spins of the corresponding states in ^{181}Ta , have been inferred from the theoretical considerations outlined below.

Theoretical considerations

Since both the atomic number and the transition energies involved in the decay of the metastable state of ^{181}Ta are relatively high, the assumptions made in the approximate calculations of internal conversion coefficients, such as those of Hebb and Nelson (1940), are not valid in this case. However, it seems possible to make an order of magnitude estimate of the coefficients involved here from the exact calculations of Taylor and Mott (1932) which are applicable to internal conversion in RaC .

The results of Taylor and Mott are given in terms of the branching ratio $\alpha/(1-\alpha)$. For the 0.5 mev. transition we find our observed value of the branching ratio to be 80% greater than that calculated for an electric quadrupole transition. For an electric dipole transition the calculated ratio is four times smaller, and for a magnetic dipole transition three times larger, than the observed value. Considering the experimental uncertainties, including self-absorption, the possibility of ejection of photoelectrons from the source, and the difference in

* Since the completion of the work described here, Cork, Shreffler and Fowler (*Phys. Rev.*, 1947, **72**, 1209) have reported from the measurement of conversion electron energies in a β -ray spectrograph that γ -rays of energies 133, 345 and 478 kev. are present in the decay of ^{181}Hf . It is difficult to reconcile these energy values with the results reported here.

atomic number, there seems to be reasonable agreement between the observed value of the branching ratio and the value calculated by Taylor and Mott for electric quadrupole radiation. Thus, although there might be a small admixture of magnetic dipole radiation, the majority of the 0.5 Mev. transitions would seem to give rise to electric quadrupole radiation. Since it is known that the lifetime of a state preceding such a transition is about 10^{-12} seconds, the state preceding the 0.5 Mev. transition cannot be metastable, but must be the lower excited state of ^{181}Ta . The spin of the ground state is $7/2$ and the multipole order of the transition is either 0 or 2. The lower excited level must therefore have the same parity as the ground state and a spin equal to $7/2$, $5/2$ or $3/2$ units of $\hbar$.

As regards the 0.2 Mev. radiation which is emitted in the decay of the 20-microsecond metastable state, an indication of the spin change can be obtained from the rough energy-lifetime relation which has been deduced from the liquid-drop model of the nucleus (Berthelot 1944). Taking into account the correction factor for internal conversion, we find the calculated value for the half-life is of the order of the observed value (the latter is seven times larger) if we assume electric octopole or magnetic quadrupole radiation to be emitted in the transition. The observed value of the branching ratio, 1.5 in this case, is, on the other hand, less than half the value calculated by Fisk and Taylor (1934) for magnetic quadrupole radiation ($Z=84$). Although exact numerical calculations do not exist for electric multipole transitions of order higher than 2, it seems likely, from a rough estimate of the variation in the value of the branching ratio with multipole order, that most of the 0.2 Mev. radiation is of an electric octopole nature. The metastable state will then have a parity opposite to that of the lower excited state, while the spins will be $1/2$ and $7/2$ respectively.

ACKNOWLEDGMENTS

The authors wish to express their appreciation to Professors M. L. Oliphant, F.R.S. and P. B. Moon, F.R.S. for their sustained interest in this work, to Dr. F. E. Whitmore for help with the experimental work, to the Director of the Atomic Energy Research Establishment for the irradiation of the hafnium sample, and to the Superintendent of the Atomic Energy Division at the Telecommunications Research Establishment for help with electronic equipment for this and other work.

REFERENCES

- BERTHELOT, A., 1944, *Cahiers de Physique*, No. 19, March.
 BITTENCOURT, P. T., and GOLDBABER, M., 1946, *Phys. Rev.*, **70**, 780.
 DEBENEDETTI, S., and MCGOWAN, F. K., 1946, *Phys. Rev.*, **70**, 569.
 DUNWORTH, J. V., 1939, *Nature, Lond.*, **144**, 152.
 FISK, J. B., and TAYLOR, H. M., 1934, *Proc. Roy. Soc. A*, **146**, 178.
 HEBB, M. H., and NELSON, E., 1940, *Phys. Rev.*, **58**, 486.
 MADANSKY, L., and WIEDENBECK, M. L., 1947, *Phys. Rev.*, **72**, 185.
 ROTBLAT, J., 1941, *Proc. Roy. Soc. A*, **177**, 260.
 SEREN, L., FRIEDLANDER, H. N., and TURKEL, S. H., 1947, *Phys. Rev.*, **72**, 888.
 TAYLOR, H. M., and MOTT, N. F., 1932, *Proc. Roy. Soc. A*, **138**, 665.

The Penetrating Particles in Cosmic-Ray Showers: I. Heavily-ionizing Particles.

BY G. D. ROCHESTER AND C. C. BUTLER
The Physical Laboratories, University of Manchester

Communicated by P. M. S. Blackett ; MS. received 7 May 1948

ABSTRACT. It is shown that whilst almost all the heavily-ionizing particles occurring in cosmic-ray showers are protons, a small number are slow mesons. Examples of the local creation of mesons in explosive-type showers are found and are identified as mainly μ -mesons. The results are consistent with the recent discovery of Gardner and Lattes that μ -mesons may be created directly and are not always the disintegration products of π -mesons as suggested by Marshak and Bethe.

§ 1. INTRODUCTION

AN account is given of an investigation of the heavily-ionizing particles which occur in cosmic-ray showers; five examples of the local creation of mesons in explosive-type showers are found (for a note on one of these showers see Rochester, Butler and Runcorn 1947). The mesons are identified as mainly μ -mesons by the measurement of their momenta and the visual estimation of the density of ionization along their tracks; the ionization has been estimated by using the protons as a reference scale.

§ 2. THE EXPERIMENTAL ARRANGEMENT

Two arrangements, designated F and P, figures 1 and 2, were used to trigger the cloud chamber. Neither arrangement was rigidly selective for any particular type

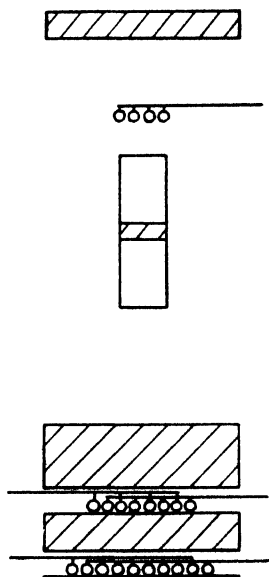


Figure 1. Arrangement F.

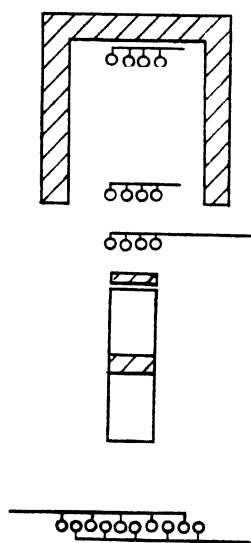


Figure 2. Arrangement P.

of shower, and even though the counters were connected in fivefold coincidence to reduce the rate of random coincidences, the cloud chamber could be set off by a shower consisting of only two particles. All the counters were 2 cm. in diameter; those above the chamber were 30 cm. long and those below 50 cm. long. In order to identify the showers in the cloud chamber associated with extensive air showers a tray of 14 counters connected in parallel, with effective area 1700 cm², was placed 1 m. from the chamber. An indicator lamp at the cloud chamber showed whether one or more ionizing particles crossed the extension tray within 2 μ sec. of the master pulse F or P. This tray was, unfortunately, in operation only during part of the present experiment. The running times and the rates for the different arrangements are shown in table 1. The data given under arrangement P are the combined results from several slightly different arrangements. The rate of associated extensive showers was approximately 2.5 showers per hour.

Table 1

Counter arrangement	Number of photographs	Time (hrs) *	(a)	(b)
F	745	378	41	25
P	4664	830	191	76

* Corrected for resetting time.

(a) No. of photographs with more than 4 particles in chamber.

(b) " " " " " " " 9 " " "

The cloud chamber was 30 cm. in diameter and 9 cm. in depth. It was filled to a pressure of 1.5 atmosphere with a gas mixture consisting of 80% argon and 20% oxygen. Across the chamber was a lead plate 30.5 mm. thick, faced above and below with 1.8 mm. brass reflecting plates. Photographs were taken at field strengths of approximately 3500 gauss and 7000 gauss using the magnet built by Blackett (1936). Illumination was provided by two Siemens S.F.4 flash discharge tubes each 10 in. long by $\frac{3}{4}$ in. wide, with a narrow guide tube inside the main tube. Each tube was triggered by a 20-kv. pulse from a magneto coil and flashed by the discharge of a bank of condensers, capacity 200 μ F., charged to 1800 v. The pulse was delayed 0.05 sec. after the chamber expansion in order to allow the drops to increase sufficiently in size to be photographed. The best film was found to be Kodak R55, a moderately fine-grain orthochromatic recording film.

The chamber was photographed by a stereoscopic pair of corrected Taylor, Taylor, Hobson (Cooke) F/2 lenses (used at F/4), of focal lengths 35 mm.

§ 3. MEASUREMENT OF CURVATURE AND ACCURACY OF ENERGY DETERMINATION

Most of the curvatures of the tracks discussed in this paper were determined by direct measurement on a travelling microscope fitted with a Zeiss micrometer eyepiece, but some were measured on the curvature-compensating machine (Blackett 1937). The results for good tracks measured by different observers agree to within 10%. A good track is defined as one which is longer than 6 cm. in the chamber and is free from obvious distortion. A more detailed discussion of the measurement of the curvatures of the tracks of high-energy particles will be given in paper II, but it may be stated here that chamber errors, which are principally due to a gas distortion, limit the maximum detectable momentum in these experiments to 6×10^9 ev/c. in a magnetic field of 7500 gauss.

The magnetic field increased by about 5% from the front to the back of the chamber and fell off by about 10% from the centre to the circumference. In most cases the correction for the variation in the field along a track was so small that it was not applied.

The exact positions of the tracks in the chamber and the magnifications were found by reprojecting the photographs through the same lenses. The momenta were derived, where possible, from measurements on both stereoscopic photographs.

§ 4. SLOW PROTONS

On 42 of the showers 61 heavily-ionizing particles have been observed, and of these 29 have been accurately measured. The ionization density along the track of each has been estimated visually by several independent observers, and the results are plotted in figure 3, following a procedure used by Leprince-Ringuet, Lh  ritier and Richard-Foy (1946). The estimated error in the ionization is taken as the range of the independent estimates.

It has not been found possible to estimate an ionization greater than fifteen times the minimum. The theoretical ionization curves calculated for masses of $200 m_e$, $320 m_e$ and $1837 m_e$ are also shown, and it is at once apparent that the slow particles fall into two main groups, one around the proton curve and the other around the curve for mass $200 m_e$. The

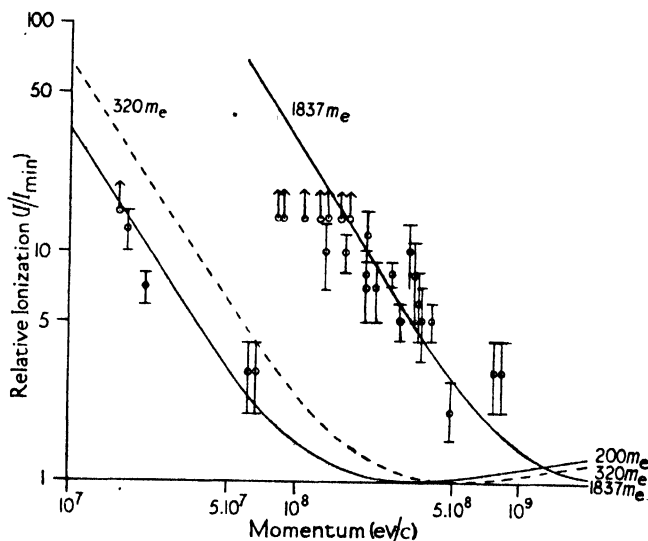


Figure 3.

spectrum of the twenty-four measured protons is given in table 2. The other slow particles which cannot be measured are probably protons of momenta 10^8 ev/c. or less. If they are added to the table it would appear that there is a rather rapid decrease in the number of protons with increasing momentum. With one exception all the protons are positive if coming downwards; the exception may well be a positive proton moving upwards.

Table 2. Spectrum of slow protons

Momentum range $\Delta p \times 10^8$ ev/c.	< 1	1-2	2-4	4-6	6-8
Number of protons in range Δp	2	8	10	2	2

Forty of the protons are associated with explosive-type showers (or "stars") and sixteen with electron-cascade showers. A typical star, in which most of the products of the disintegration have been projected forwards, is shown in the plate, photograph (1), and an example of a more complex shower in photograph (2). In this shower a particle (m) with a momentum of 4.5×10^8 ev/c. produces a positive heavily-ionizing particle (p) of momentum 3.2×10^8 ev/c. which is.

probably a proton. This photograph may be an example of a nuclear explosion induced by a σ -meson, a process already well known from the work of Powell and Occhialini.

§ 5. SLOW MESONS

(i) *Experimental data*

Data for four of the slow mesons obtained in these experiments are given in table 3 and photographs of these particles are reproduced in the plate, photographs (3), (4), (5) and (6). Data and a photograph of a fifth slow meson (designated meson (7)) have already been published (Rochester, Butler and Runcorn 1947).

Table 3

Photograph number in plate	(3)	(4)	(5)	(6)
Sign	Positive		Negative	Negative
Magnetic field (gauss)	7100	7100	7250	6500
Mean rad. of curvature in plane of chamber (cm.)	25.0	5.8	29.0	9.4
Calc. rad. of curvature due to scattering (a)	280	54	200	106
" " " " (b)	140	27	100	53
Momentum (ev/c.)	6.9×10^7	1.7×10^7	6.3×10^7	2.2×10^7
Kinetic energy (ev.)	2.4×10^7	1.4×10^6	2.0×10^7	2.4×10^6
Estimated $I/I_{\min.}$	2-4	15	2-4	10-15
Calc. $I/I_{\min.}$: Electron (m_e)	1.5	1.4	1.3	1.4
Meson ($100 m_e$)	1.2	5.6	1.2	3.8
Meson ($200 m_e$)	2.1	16.2	2.3	10.9
Meson ($320 m_e$)	3.9	32.4	4.5	22.3
(a) Williams (1939)	(b) Bethe (1946).			

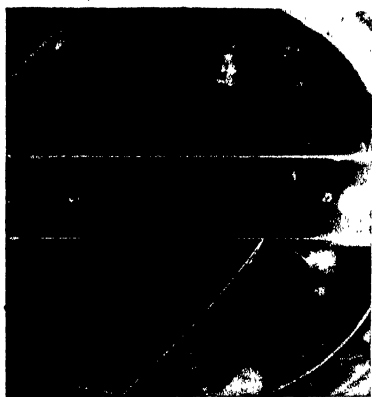
This photograph (taken with arrangement F) was a clear example of the production of a negative meson in an explosive-type shower in the lead plate. Mesons (3), (4) and (5) were all photographed with arrangement P. In the case of meson (6) there was *no* lead above the chamber and the extension was set off by the shower.

Where possible, the sign of the particle has been determined by making a reasonable assumption as to the direction of the particle, and the choice is confirmed in several cases by other evidence. All these slow particles make large angles with the vertical. In three cases, (3), (5) and (7), the mesons arise in nuclear explosions in the lead plate. The explosion associated with meson (3) cannot be seen clearly on the reproduction, for the main part is out of the light at the front of the chamber. It can, however, be seen faintly on the original negative.

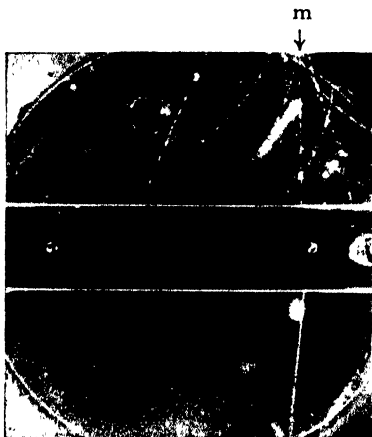
Because of the large magnetic field, the possible error in the radius of curvature of the track due to scattering is relatively small. It has been calculated both from Williams' and from Bethe's formulae, and the results are given in the table.

The estimated ionization is given in line 8, and in line 9 are given the calculated relative ionizations for particles of electronic mass (m_e) and masses $100 m_e$, $200 m_e$ and $320 m_e$. It is seen from the table and from figure 3 that in most cases reasonable agreement between estimated and calculated ionizations is obtained by assuming a mass of approximately $200 m_e$, although in three of the cases a higher value cannot be excluded.

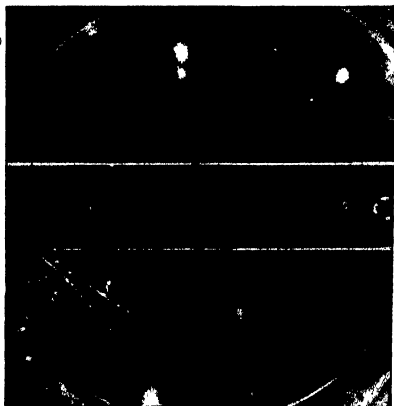
Confirmation is obtained in the case of particle (6) by the knock-on electron indicated by the arrow "k" in the photograph. Another electron, which on the reproduction appears to be a second knock-on electron, is not in the same plane as the meson. The kinetic energy of the knock-on electron estimated from



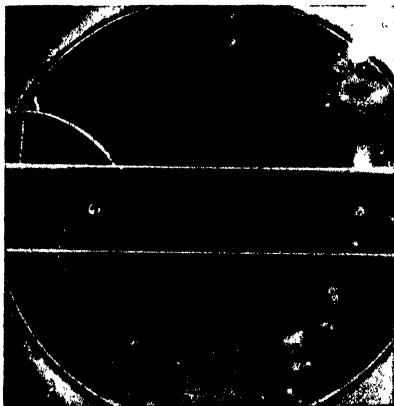
(1)



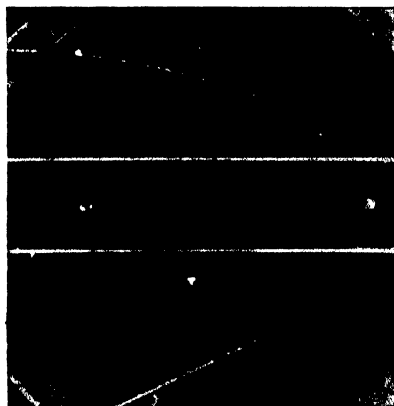
(2)



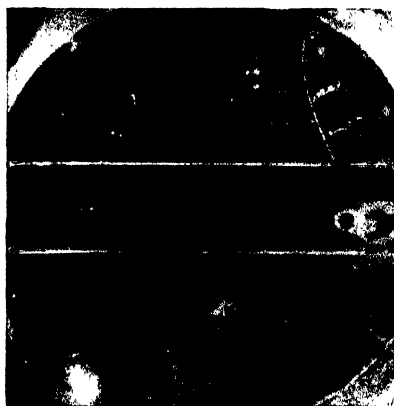
(3)



(4)



(5)



(6)

(See back of plate for description.)

DESCRIPTION OF PLATE

Photograph (1). An example of a typical star created in the lead plate by non-ionizing radiation. All the observed particles are positive and they are identified as protons. The heavily-ionizing particles have kinetic energies of a few mev. and the lightly-ionizing particles kinetic energies of approximately 100 mev. The effects of gas scattering (*not* chamber distortion) can be observed at the ends of two of the proton tracks. One of these particles stops in the chamber.

$H=7200$ gauss.

Photograph (2). An example of the production of a proton (p) of momentum 3.2×10^8 ev/c. by a positive, lightly-ionizing particle (m) of momentum 4.5×10^8 ev/c., which is associated with an electron shower.

$H=7100$ gauss.

Photograph (3). An example of the production of two protons and a positive meson of momentum 6.9×10^7 ev/c. in an explosive-type shower which is close to the front of the chamber.

$H=7100$ gauss.

Photograph (4). A very slow meson of momentum 1.7×10^7 ev/c. There is slight electrostatic field doubling at the lead plate.

$H=7100$ gauss.

Photograph (5). An example of the creation in the lead plate of a negative meson of momentum 6.3×10^7 ev/c. and a proton of 1.3×10^8 ev/c.

$H=7250$ gauss.

Photograph (6). A slow meson associated with an extensive air shower. This meson, which is interpreted as a μ -meson, has a momentum of 2.2×10^7 ev/c. (kinetic energy 2.4×10^8 ev.) and is negative if coming downwards. It comes across the chamber out of the piston and may therefore have been created locally. The direction and identification are confirmed by the knock-on electron (k). The other electron, which appears to be a second knock-on, can be seen stereoscopically to cross the meson track towards the front of the chamber.

$H=6500$ gauss.

its range is 60 000 ev. (Nuttall and Williams 1926). The calculated energies of knock-on electrons from particles of masses $200 m_e$, $320 m_e$ and $1837 m_e$ for "head-on" collisions are 46000 ev., 18000 ev. and 600 ev. Thus a mass of $200 m_e$ is consistent with the measured value of the energy of the electron although a somewhat lower mass cannot be excluded in view of the fact that the angle between the electron track and the meson track cannot be measured.

(ii) Discussion

These results have a bearing on the important problem of the creation of the meson. Three of the mesons are observed to be created in the lead plates and in showers of the nuclear-explosive type and not in electron-cascade showers. W. E. Powell (1946) and Fretter and Hazen (1946) also found that slow mesons appeared predominantly in explosive-type showers and is consistent with the recent discovery of the artificial production of mesons by the Berkeley cyclotron (Gardner and Lattes 1948).

The relation of the mesons observed in these photographs to the new types discovered by C. F. Powell, Occhialini and their co-workers will now be discussed. It seems plausible to assume that the Powell μ -meson is to be identified with the ordinary cosmic-ray meson found at sea-level and that the mass of the π -meson is approximately $320 m_e$ (Occhialini and Powell 1948). The evidence of Powell and his co-workers is that these heavy mesons form a large fraction of all slow mesons found at moderately high altitudes and that the π -meson decays into a μ -meson with a kinetic energy of 4 Mev. Since some of the mesons found in the present experiment seem to have masses of approximately $200 m_e$, the following assumptions may be made : (a) the observed mesons are mainly μ -mesons created directly in nuclear explosions, (b) the observed mesons are μ -mesons arising from the disintegration of π -mesons within the lead plate. If the latter is the correct interpretation, the lifetime must be exceedingly small, certainly less than 10^{-10} sec. Recent evidence from the artificial production of mesons, however, suggests that the lifetime of the π -meson is very much longer than 10^{-10} sec., perhaps of the order of 10^{-8} or 10^{-9} sec. Again, in two of the cases the kinetic energies of the observed mesons are much greater than 4 Mev. ; thus these particles could not have been created by the disintegration of π -mesons at rest. It would appear therefore that assumption (a) represents the more likely explanation of the present photographs. This result is consistent with preliminary reports from Berkeley which indicate the direct creation of μ -mesons.

The meson (6) is of interest because it is connected with an extensive air shower. This meson comes from the piston at the back of the chamber, and it may therefore have been created locally. The photograph was selected from some 500 photographs of extensive air showers, the data for which will be presented in a later paper. The photograph is a clear indication that ordinary cosmic-ray mesons are found in extensive air showers, but it gives no information as to whether such mesons are created locally or come from the air.

ACKNOWLEDGMENTS

It gives us great pleasure to express our sincere thanks to Professor P. M. S. Blackett, F.R.S., for the excellent facilities which he has given us and for the keen interest he has taken in the investigation. We are also indebted to Dr. S. M. Mitra for help in taking two of the photographs used in this paper.

REFERENCES

- BETHE, H. A., 1946, *Phys. Rev.*, **70**, 821.
 BLACKETT, P. M. S., 1936, *Proc. Roy. Soc. A*, **154**, 564 ; 1937, *Ibid.*, **159**, 1.
 FRETTER, W. B., and HAZEN, W. E., 1946, *Phys. Rev.*, **70**, 230.
 GARDNER, E., and LATTES, C. M. G., 1948, *Science*, March 12.
 LEPRINCE-RINGUET, L., LHERITIER, C., and RICHARD-FOY, R., 1946, *J. Phys. Radium*, sér. 8, **7**, 69.
 MARSHAK, R. E., and BETHE, H. A., 1947, *Phys. Rev.*, **72**, 506.
 NUTTALL, J. M., and WILLIAMS, E. J., 1926, *Phil. Mag.*, ser. 7, **2**, 1109.
 OCCHIALINI, G. P. S., and POWELL, C. F., 1948, *Nature, Lond.*, **161**, 551.
 POWELL, W. E., 1946, *Phys. Rev.*, **69**, 385.
 ROCHESTER, G. D., BUTLER, C. C., and RUNCORN, S. K., 1947, *Nature, Lond.*, **159**, 227.
 WILLIAMS, E. J., 1939, *Proc. Roy. Soc. A*, **169**, 531.

Analysis of Spurious Counts in Geiger Counters

By J. L. PUTMAN

Atomic Energy Research Establishment, Harwell

Communicated by J. D. Cockcroft; MS. received 25 March 1948

ABSTRACT. A method is described by which the intervals between successive pulses may be divided into size classes. The rate of arrival of intervals whose length is between t and $(t + \delta t)$ is measured, where t is variable and δt is a preset period, small compared with t .

The method is applied to the study of Geiger counter pulses for argon-alcohol counters. By examining the deviation of the pulse interval distribution from random laws, the number and time distribution of spurious counts is deduced for various potentials applied to the counter. It is found that almost all the spurious counts occur during the recovery period of the counter. They may be eliminated by the use of an external quenching circuit which applies an artificial dead time of the order of 400 microseconds to the counter.

§ 1. INTRODUCTION

WHEN a Geiger counter is used to record the passage of ionizing particles, a number of pulses are delivered by the counter which do not correspond directly to the passage of particles, but which are indistinguishable from the pulses which do correspond to the particles to be counted. These are usually called spurious counts, and an allowance must be made for them in accurate measurements with a Geiger counter.

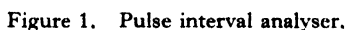
Spurious counts have been defined by Korff (1946) as those "caused by any agency whatever other than the entity which it is desired to detect, or the normal contamination, or cosmic-ray background". If we accept this definition, we must include as spurious counts those which are caused by any unwanted external agency, such as the impact of visible light on the counter.

In the present paper, we shall confine our attention to those counts which are not caused directly by any external agency, but which are generated inside the counter as a result of earlier discharges.

When a Geiger counter is used with a radioactive source of long half-life, the arrival of ionizing particles in the counter is purely random, and the time spacing between successive particles follows well-known laws (Marsden and Barratt 1911). In particular, the probability of an interval of length between t and $(t + \delta t)$ is

$$P(t) = fe^{-ft} \cdot \delta t, \quad \dots\dots(1)$$

The distribution of time intervals between Geiger counter pulses has already been used as an indication of the time delays of spurious counts, notably by Medicus (1936). Interval selectors have also been designed by Roberts (1940).



Curran and Rae (1947) used a system of delayed coincidences to determine the rate of arrival of pulse-spacings of discrete lengths. It should be noted, however, that the spacings they measure are not necessarily between successive pulses. Consequently the rate of arrival of intervals of a given length, as measured by this method, is a function which depends on the dead-time of the counter and can only be expressed in a simple form when certain approximations are made.

21

small compared with t . With this equipment, the distribution of spacings between successive pulses is analysed. This distribution is directly comparable with equation (1), and the effects of dead-time and of spurious counts upon the random distribution can be shown explicitly.

§ 2. THE PULSE INTERVAL ANALYSER

A simplified form of the circuit used in the pulse interval analyser is shown in figure 1. Pulses from the Geiger counter which is to be examined are fed through an amplifier and limited so that a positive pulse of fixed size and short duration is fed to the grid of V_1 for each pulse delivered by the Geiger counter.

This pulse causes V_1 to conduct and to discharge the condenser C_1 . At the same time the left-hand side of V_2 stops conducting and the potential of the left-hand anode A rises sharply.

The condenser C_1 then recharges through the leak R, and at a time t after the arrival of the pulse, the left-hand grid of V_2 rises to a conductive level and the potential of A falls sharply to its original level. If a subsequent pulse is delivered by the Geiger counter before the time t has elapsed, C_1 is again discharged and the recovery process begins again. The potential of the point A cannot fall to its original value until a pulse-free interval t has elapsed. Figure 2(a), (b) and (c) shows the sequence of events. In figure 2(c) the trailing edge of a square wave at A marks a time t after the arrival of the last pulse. The leading edge corresponds to the first pulse to arrive after a pulse-free interval t has elapsed.

The output from A is reversed in phase by V_3 , the output from which is then differentiated and fed to the grid of V_4 (figure 2(d)). Thus a positive pulse, corresponding to the trailing edge of the square wave at A, is used to trigger the flip-flop V_4V_5 which delivers a short positive pulse of length δt to the suppressor V_8 (figure 2(e)).

The suppressor of V_8 is thus raised to a conductive level for the period δt immediately after a pulse-free interval t has elapsed.

At the same time the output from A is differentiated and applied to the control grid of V_8 , so that the leading edges of the square waves at A give rise to positive pulses (figure 2(f)) which raise the grid of V_8 to a conductive level. If this happens during the period δt in which the suppressor potential is also raised, V_8 conducts and registers an output pulse (figure 2(g)).

Thus an output pulse is only obtained when a pulse occurs in the period between t and $(t + \delta t)$ after the previous one, i.e. when a pulse-free interval between t and $(t + \delta t)$ occurs. The output pulse is fed to a scaler and recorder unit.

The time t is adjustable by means of the potentiometer R, and can be set at from 4 to 670 μsec . The short period δt has one of a number of values from 1 to 20 μsec . selected by switching the condenser C_2 in the flip-flop.

We would mention that the pulse interval analyser has a wider application than the study of Geiger counter pulses. It could be used, for example, to determine the random or systematic nature of electronic noise pulses, or for the determination of short radioactive half-lives.

§ 3. PULSE INTERVAL DISTRIBUTIONS

It was found convenient to measure the rate of arrival of periods of given lengths, using the pulse interval analyser, and to compare these with the distribution for random events :

$$r(t) = f^2 e^{-ft} \delta t, \quad \dots\dots(2)$$

which is derivable from equation (1) by multiplying by f , the mean rate of arrival of particles.

A few counters of different types were used for this analysis. All were of the self-quenching type, being filled with a 9 : 1 mixture of argon and alcohol.

The interval tolerance δt was kept constant at a value of $8.3 \mu\text{sec}$. whilst t was set (in the analyser) at a series of values from 50 to $650 \mu\text{sec}$. The rate of arrival of pulse intervals was recorded for each setting of t , to a statistical accuracy of 3%. This rate was plotted against the lower length t of the intervals. The total counting rate was also recorded with a rate-of-count meter of an integrating type. The experiments were repeated using a number of applied potentials to each counter.

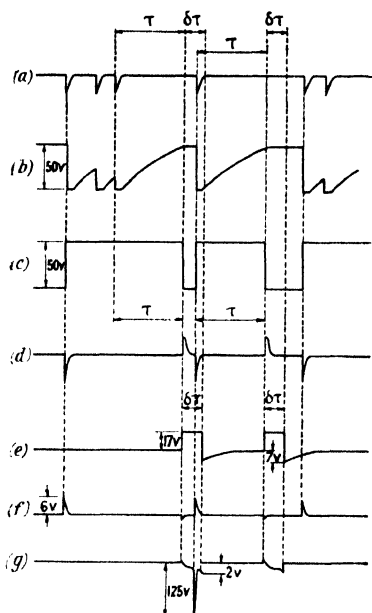


Figure 2.

Waveforms from pulse interval analyser for "Analysis of spurious counts in Geiger counters".

- (a) Geiger counter pulses.
- (b) Anode of V1. (See figure 1)
- (c) Anode A of V2. (See figure 1)
- (d) Control of V4. (See figure 1)
- (e) Suppressor of V8. (See figure 1)
- (f) Control of V8. (See figure 1)
- (g) Output anode of V8. (See figure 1)

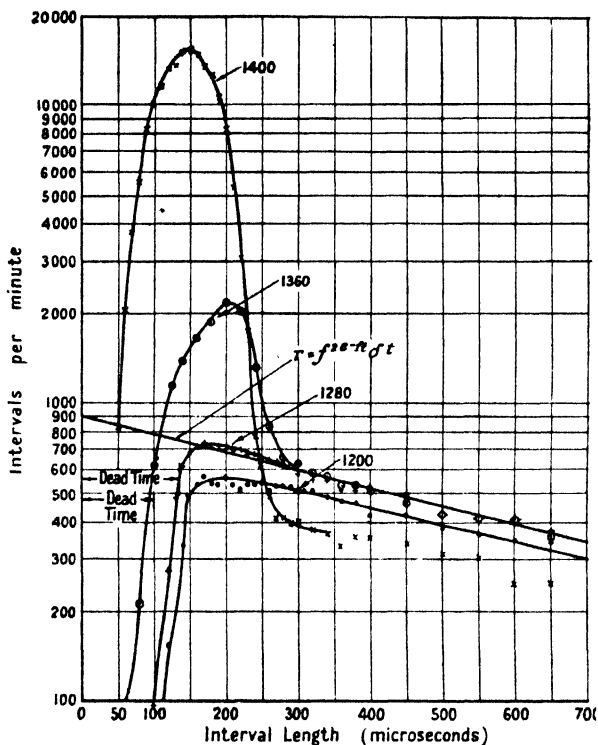


Figure 3. Distribution of intervals between Geiger counter pulses.

Note: The rate of arrival of pulse intervals whose length is between t and $(t + \delta t)$ is plotted as a function of t . $\delta t = 8.3 \mu\text{sec}$ throughout.

Figure 3 shows the distributions of intervals measured for a commercial end-window type beta counter. The rate of arrival of intervals of lengths between t and $(t + \delta t)$ is plotted on a logarithmic scale against a linear scale of interval lengths t .

For the lower applied potentials, the rate of arrival rose steeply with increasing t to a maximum value, from which it fell exponentially as t was further increased. The dead time of the counter, which was measured separately, is indicated in figure 3 for two applied potentials. The occurrence of some intervals which were

shorter than the normal dead-time corresponds to pulses arriving after the shorter dead-times which followed pulses in a partially recovered counter. (A separate investigation showed that counts occurring during the recovery time were in fact followed by dead-times shorter than normal, a result which was to be expected since the partially recovered pulses are accompanied by space charges smaller than normal.)

Since the interval rate rose to its maximum very shortly after the dead-time, it would appear that very few, if any, counts were lost by recombination of primary electrons with the outward moving positive ions in the space charge.

The exponential fall in the interval rate, as t was increased, was in conformity with equation (2). Thus for the lower applied potentials the pulse distribution was purely random except where it was limited by the dead-time of the counter. The true distribution of intervals between particles passing through the counter was deduced by extrapolating the distribution from longer intervals back to $t=0$. The process is analogous to the dead-time correction normally made in counting experiments. The time δt was also deduced accurately by comparing the intercept on the axis ($t=0$) with the measured total counting rate f .

For higher applied potentials (see figure 3) a peak appeared in the distribution curve near $t=200\ \mu\text{sec}$. Here the rate of arrival of intervals was greater than that to be expected from a random distribution of pulses. Excessive rates were found to extend from the dead-time to the recovery time of the counter. These critical times were measured separately. For values of t greater than the recovery time the interval rates decreased exponentially as t was increased, and the interval distribution followed closely that measured at lower operating potentials.

Evidently the additional intervals in excess of the random number corresponded to spurious counts generated internally in the counter on the arrival of positive ions at the cathode. The mechanism has been described by Korff and Present (1944). The increase in the number of spurious counts of this kind accounted very largely for the slope in the plateau of the counter. Good correlation was obtained between the total number of counts, as integrated under the distribution curve, and the total measured counting rate.

A further contribution to plateau slope arose from the slight increase of general level of the random count rates as the applied potential was increased over the lower ranges. This was considered to be due to end effects in the short counter used, which resulted in an increase in the sensitive volume of the counter as the applied potential was raised. The effect disappeared when a long thin counter was used, so that end effects became negligible. When the long counter was used, the distribution of pulse intervals became independent of applied potential for values of t greater than the recovery time for the counter, i.e. all the distribution curves coincided for the higher values of t .

At very high potentials, the general level of random intervals fell below the normal level because of the increased loss in dead times associated with the large number of spurious counts. This is shown in the curve for 1400 volts in figure 3. The counter was then operating well above the plateau level, a condition not likely to be met in practice.

§ 4. SUPPRESSION OF SPURIOUS COUNTS

Since all the spurious counts observed in the above experiments occurred within the recovery time, it was decided to render the counter inoperative after

each count for a period slightly greater than the recovery time, in order to examine the flattening of the plateau produced by this means.

This was achieved by the use of an improved form of Getting's (1938) quenching circuit designed by Cooke-Yarborough. This constituted a flip-flop which was triggered by incoming pulses and which caused the potential applied to the counter to fall by some 270 volts for a period of the order of $400\ \mu\text{sec.}$ and then to revert quickly to its normal level. Thus any electrons produced in the counter during the $400\ \mu\text{sec.}$ period would find the counter reduced in amplification factor below the Geiger-Müller region and would produce pulses too small to be recorded. An artificial dead-time was thus applied to the counter.

When the choice of the applied dead-time was made, an allowance was made for the increase in recovery time due to the reduced potential.

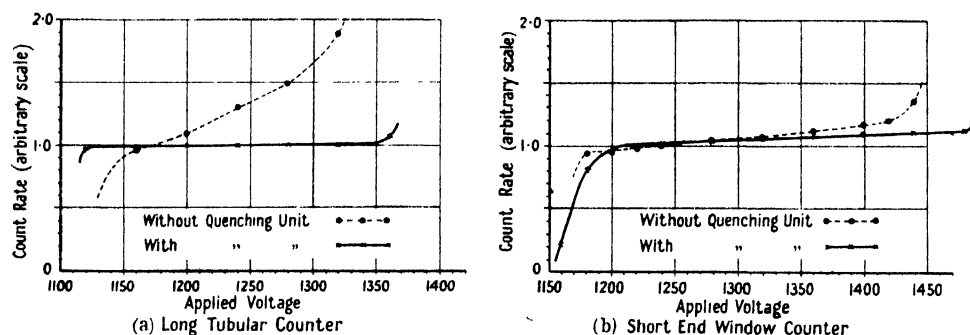


Figure 4. Effects of quenching circuit on the plateau.

The results of applying the quenching circuit are shown in figures 4(a) and 4(b) which are for long and short counters respectively. It can be seen that, with the long counter, an initially rather bad plateau became perfectly flat when the quenching circuit was used. This shows that the initial slope was almost entirely due to spurious counts. Great improvement in plateau slope was also obtained with the short counter, but a small residual slope persisted, which could be attributed to the end effects mentioned above.

The use of a quenching circuit of this kind with a self-quenched counter to remove spurious counts has a practical significance where accurate counting is required. It was found to be of particular value in eliminating variations of counting rate with temperature, which were traced to a bodily shifting of the plateau towards high operating potentials as the temperature was increased. When the plateau was flattened with an external quenching circuit of the form just described, the counting rate became largely independent of temperature.

§ 5. DISCUSSION OF RESULTS

The position of the peak in the distribution curve of figure 3 indicates very strongly that the spurious counts were mainly due to secondary electrons liberated from the cathode on the arrival of positive ions. This explanation is strengthened by the limitation of the intervals preceding spurious counts to a period equal to the recovery time of the counter.

Another possible source of spurious counts lies in the formation of negative ions, as, for example, when electrons are captured by neutral oxygen atoms (Rose and Ramsey 1941, Montgomery and Montgomery 1941 a). These heavy ions

move much more slowly towards the anode of the counter than do free electrons in the avalanche process. The captured electrons are released in the high fields near the anode wire of the counter and give rise to delayed spurious counts. The time delays involved are of the same order as those for spurious counts caused by cathode secondaries. The presence of a large number of spurious counts due to electron capture, as well as those due to cathode secondaries, would be expected to cause a double maximum in the distribution curve for pulse intervals. One counter which was suspected of having an air leak was tested and some indication of a double maximum was found. The results, however, were inconclusive.

The spurious counts shown in figure 3 are distributed over a wide range of pulse intervals t from the dead-time to the recovery time of the counter. This result indicates that positive ions are arriving in large numbers at the cathode throughout this period. The positive space charge is therefore quite diffuse by the time it reaches the cathode and not concentrated in a thin shell as has been assumed (Montgomery and Montgomery 1941 b, Stever 1942). The result is not surprising, since the forces of mutual repulsion acting on such a shell would be of the same order as the applied fields in the counter, so that spreading would take place as the space charge moved outwards towards the cathode.

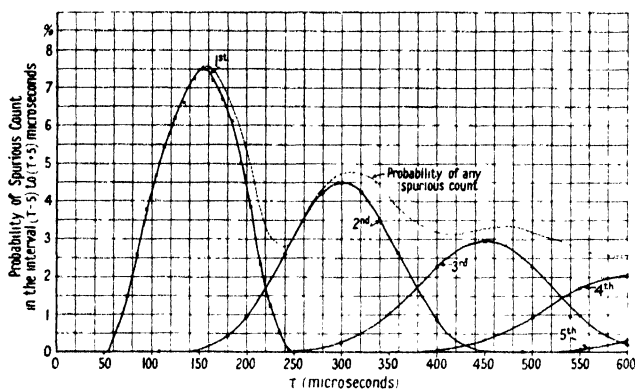


Figure 5. Spurious counts of higher orders.
(Deduced from distribution of figure 3 for 1400 volts.)

These results, which were obtained in 1946*, are in good agreement with subsequent measurements by Curran and Rae (1947) using a different method.

It should be noted that a spurious count is almost as likely to produce a further spurious count as is a true count. A second spurious count would be related to the first by a time delay of the same magnitude as though it followed a true count. Hence in figure 3 the excess rates recorded correspond to spurious counts of the first, second, third order, and so on.

If the probability of a count being followed by a spurious count is P_1 , then if the true count rate is f , the rate of arrival of first spurious count is $S_1 = fP_1$. Now, if spurious counts are assumed for simplicity to have the same chance as true counts of being followed by spurious counts, then the rate of arrival of second spurious counts is $S_2 = S_1P_1 = fP_1^2$, and proceeding thus, the rate of arrival of n th spurious

* The Pulse Interval Analyser described in § 2 was used to demonstrate these results at the Physical Society Exhibition of Scientific Instruments and Apparatus, April 1947.

counts is $S_n = fp_1^n$. Thus, if p is the total probability of a spurious count, we have $p = \sum_{n=1}^{\infty} p_1^n = p_1/(1 - p_1)$, so that the total spurious count rate is $s = fp = fp_1/(1 - p_1)$.

Using this relation, the distributions of spurious counts up to the 5th order was deduced from the curve for 1400 volts in figure 3. These distributions are shown in figure 5. A curve is superimposed which shows the total probability of occurrence of a spurious count at a given time after a true count.

The assumption of equal probability for all spurious counts is not strictly justified, since those spurious counts which occur after the dead-time are associated with smaller space charges than normal and consequently have a reduced chance of producing a further spurious count. However, the method of assessing the distribution of figure 5 is probably justified as a first approximation.

It is certainly true that when the number of spurious counts becomes large, the probability of higher order counts becomes appreciable. For this reason a quenching circuit of the form described in §4 is to be preferred to an artificial paralysis in the recording circuits as a method of removing spurious counts. The quenching circuit eliminates all the spurious counts by preventing the occurrence of those of the first order. The paralysis method rejects the first-order spurious counts but allows those of higher order to be recorded.

§ 6. CONCLUSIONS

We conclude that the major cause of plateau slope in argon-alcohol filled counters is the occurrence of spurious counts during the recovery period. These are almost certainly due to secondary electrons liberated by positive ions arriving at the cathode of the counter. By the use of an external quenching circuit which applies an artificial dead-time greater than the recovery time of the counter, very flat plateaux can be obtained, especially for long thin counters.

ACKNOWLEDGMENTS

The author wishes to thank Mr. E. H. Cooke-Yarborough of A.E.R.E. for his interest in the work and for his help in the design of electronic circuits. The work was a part of the research programme of A.E.R.E. Harwell and the results were published as an official memorandum in February 1947. Acknowledgment is made to the Director, A.E.R.E. and the Ministry of Supply for permission to publish this paper.

REFERENCES

- CURRAN, S. C., and RAE, E. R., 1947, *Rev. Sci. Instrum.*, **18**, 871.
 DRISCOLL, R. L., HODGE, M. W., and RUARK, A., 1940, *Rev. Sci. Instrum.*, **11**, 241.
 GETTING, I. A., 1938, *Phys. Rev.*, **53**, 103.
 KORFF, S. A., 1946, *Electron and Nuclear Counters* (New York: Van Nostrand), p. 16.
 KORFF, S. A., and PRESENT, R. D., 1944, *Phys. Rev.*, **65**, 274.
 MARSDEN, E., and BARRATT, T., 1911, *Proc. Phys. Soc.*, **23**, 367.
 MEDICUS, A., 1936, *Z. Phys.*, **103**, 76.
 MONTGOMERY, C. G., and MONTGOMERY, D. D., 1941 a, *Phys. Rev.*, **59**, 1045; 1941 b, *J. Franklin Inst.*, **231**, 509.
 ROBERTS, A., 1940, *Phys. Rev.*, **57**, 564.
 ROSE, M. E., and RAMSEY, W. E., 1941, *Phys. Rev.*, **59**, 691.
 STEVER, H. G., 1942, *Phys. Rev.*, **61**, 38.

The Mechanism of the Reversible Thermal After-effects in Glass at Ordinary Temperatures

By G. O. JONES

The Clarendon Laboratory, Oxford

MS. received 6 February 1948

ABSTRACT. An explanation is suggested for the reversible thermal after-effects shown by glass, which are responsible for the zero-depression in mercury-in-glass thermometers. A rough relation is established between these effects and the delayed elastic effects shown under ordinary mechanical stress. It is assumed that the thermal after-effects are delayed elastic strains resulting from stresses between regions of slightly differing composition, the origin of the stresses being the differing contractions of these regions during cooling from the annealing temperature. The model offers explanations for some detailed features of the behaviour, requiring inhomogeneities corresponding to differences of only a few per cent in the proportions of the main constituents to account for the observed effects.

§ 1. INTRODUCTION

IT was shown by Weidmann that there was a connection between the delayed elastic effects (or elastic after-effects) and the thermal after-effects in glasses at room temperature, a glass which was "good elastically", that is, having a small after-effect, being also "good thermally". As far as is known to the writer, no closer connection than this has been established, and the actual mechanism of the reversible thermal after-effects is still unexplained. They must be connected with an interchange of energy between different parts of the structure, and it is this which makes the phenomena of interest as it implies a degree of atomic or group mobility. Three main types of atomic movements have to be considered, namely, those corresponding to the very gradual contraction of the structure as a whole towards the equilibrium supercooled liquid condition (often known as "stabilization"), those corresponding to viscous flow under the action of external or "frozen-in" mechanical stresses, and those corresponding to delayed elastic strains and recoveries. At ordinary temperatures only the latter movements have relatively short equilibrium or relaxation times, so it is reasonable to look to them for the explanation of the thermal after-effects.

This paper suggests a possible phenomenological relationship between the thermal and mechanical after-effects which, since the mechanical after-effects are familiar and relatively well understood, we can regard as an explanation of the thermal after-effects.

It should be mentioned that some of the earlier workers (e.g. Crafts 1882) suggested explanations in terms of the relaxation of "strains" in the glass, in general without specifying what type of "strain" was being considered or accounting for their presence. With the data then available it would not have been possible to estimate what degree of mobility could be expected in the glass.

§ 2. THE THERMAL AND ELASTIC AFTER-EFFECTS

The thermal after-effects under consideration are the reversible effects which are responsible for the temporary zero-depression in a mercury-in-glass thermometer rapidly cooled to the ice-point, and for the slowness of the approach to an

equilibrium reading when rapidly heated to a higher temperature. In both cases the effect is entirely due to the glass, which when heated or cooled does not immediately reach its new volume. The larger part of the total change in volume occurs instantaneously, but the last small fraction of the change occurs only slowly, with an equilibrium time of minutes, hours or days according to circumstances. In all reported cases the instantaneous and the delayed change in volume appear to have the same sign. The expansion of the mercury is wholly instantaneous, however, so that when rapidly cooled to 0° the mercury level passes the 0° reading and slowly rises to it; when rapidly heated to a higher temperature the mercury again passes the equilibrium reading and slowly (but not as slowly as in the former case) falls back to it. The magnitude of the delayed effect at any temperature reaches a maximum if the glass has been held at the other temperature for a long period, and this maximum is greater for large temperature differences. It will be convenient to mention here that a "depression constant" is defined for mercury-in-glass thermometers as the immediate apparent depression of the zero which occurs when the thermometer is rapidly cooled to 0° after being maintained for a long period at 100° . It varies between 0.01° and 1° for different glasses and is generally in the region of 0.05° . More detailed accounts of the effect have been given by Dickinson (1906) and by Guillaume (1889).

These short-term reversible effects are quite distinguishable from the very gradual "secular rise" in the zero which also occurs over a period of years as a result of the slow contraction of the glass towards a final volume.

The delayed elastic effects consist of a somewhat analogous slowly developing elastic strain additional to the instantaneous elastic strain which occurs on application of a mechanical stress. On removal of the stress both parts of the strain are recovered, the delayed part again being recovered only slowly with an equilibrium time of minutes or hours which does not apparently vary significantly with temperature. The total magnitudes of these effects, defined as the ratio R of the delayed strain to the instantaneous strain for a given stress, vary between 0.1% and 3% for different glasses at room temperature, and increase with increasing temperature. If sufficient time is not allowed for the full development of the delayed strain under load, the delayed recovery is correspondingly reduced. The mechanisms of these effects are fairly well understood; they can occur, for instance, in structures where regions of true solid matter enclose true liquid regions, and they probably occur in all liquids, including glass, in which regions of different relaxation times or viscosity occur in conjunction. In structures of the latter type true irreversible viscous flow also occurs under stress, but the viscosity is so high in the case of glass that we need not consider this in the explanation of these short-term effects. A fuller account of the delayed elastic effects has been given by the author elsewhere (Jones 1944).

§3. RELATION BETWEEN THE TWO TYPES OF AFTER-EFFECT

The basis for the establishment of a relation between the two kinds of after-effect is the assumption that the slow changes in volume represent the net result of a number of delayed elastic strains of differing magnitudes which occur in adjacent regions of the glass at different temperatures. The stresses which give rise to these delayed strains are the result of the differing degrees of contraction of these regions during cooling from the transformation or annealing

region. (We are not concerned with the temporary stresses which are introduced as a result of temperature gradients in the glass during rapid heating or cooling in the working range. Some authors (see for instance Hovestadt 1902) have considered that these temporary stresses might be responsible for the thermal after-effects, but Young (1890) and others have shown this to be incorrect.)

We consider an idealized structure in which the glass is equally divided into uniformly distributed small regions of two types, the physical properties within one type being uniform. Adjacent small regions are assumed to be in full mechanical and cohesive touch with each other. The "instantaneous" coefficients of linear expansion α , the magnitudes of the ratio R (as already defined), and the variation of this quantity with temperature are, however, assumed to differ for the two types.

We need not consider the variation with temperature of the instantaneous coefficient of expansion, nor of the instantaneous value of Young's modulus Y , assumed the same for both types of region. Employing appropriate superscripts to denote the values of R at temperatures T and T' , and subscripts 1 and 2 to denote the two types of region, the following quantities are defined: α_1 , α_2 , R_1 , R_2 , R'_1 , R'_2 .

For convenience we consider a system in one dimension, and the elementary model can then consist of two thin strips of glass of types 1 and 2 of unit length, stuck together along their lengths, or simply clamped together at their ends. This clamping must be imagined performed at the lowest temperature T_g^* at which ordinary mechanical stresses can be relaxed quickly in either of the two glasses, and the combination then cooled suddenly to the temperature T with no development of delayed strains of magnitudes corresponding to any intermediate temperature. (As already stated, we do not consider effects due to temporary temperature gradients. Neither need we consider the bending that would occur if such a single elementary model were cooled on its own.)

Now if $-\Delta$ is the instantaneous contraction when cooled from T_g to T , there will be compression stresses in 1 and 2 of $Y\{-\Delta - \alpha_1(T_g - T)\}$ and $Y\{-\Delta - \alpha_2(T_g - T)\}$ respectively. Since the actual stresses must be equal and opposite, we have $Y\{-2\Delta - (\alpha_1 + \alpha_2)(T_g - T)\} = 0$, so that

$$-\Delta = (\alpha_1 + \alpha_2)(T_g - T)/2. \quad \dots\dots(1)$$

The delayed elastic effects may be most conveniently considered by defining "final" values of Young's modulus as $Y/(1 + R_1)$ and $Y/(1 + R_2)$ for the two glasses at temperature T .

Then if $-\delta\Delta$ is the additional delayed contraction we find, by a similar argument,

$$-(\Delta + \delta\Delta) = (\alpha_1 + \alpha_2 + \alpha_1 R_2 + \alpha_2 R_1)(T_g - T)/(2 + R_1 + R_2). \quad \dots\dots(2)$$

Subtracting and rearranging, equations (1) and (2) lead to

$$\delta\Delta = (\alpha_1 - \alpha_2)(R_1 - R_2)(T_g - T)/2(2 + R_1 + R_2). \quad \dots\dots(3)$$

If $-\Delta'$ and $-(\Delta' + \delta\Delta')$ are the instantaneous and delayed contractions corresponding to a rapid cooling from T_g to T' , we have by a similar argument

$$-\Delta' = (\alpha_1 + \alpha_2)(T_g - T')/2, \quad \dots\dots(4)$$

$$\text{and} \quad \delta\Delta' = (\alpha_1 - \alpha_2)(R'_1 - R'_2)(T_g - T')/2(2 + R'_1 + R'_2). \quad \dots\dots(5)$$

* T_g is of course the conventional symbol for the transformation temperature, which is the temperature at which the equilibrium times for the relaxation of mechanical stresses under constant strain are of the order of seconds or minutes.

It is reasonable to suppose that in a rapid change from equilibrium at T to the temperature T' , the instantaneous and delayed expansions are given by $(\Delta' - \Delta)$ and $(\delta\Delta' - \delta\Delta)$ respectively. Defining instantaneous and final expansion coefficients $\bar{\alpha}$ and $(\bar{\alpha} + \delta\bar{\alpha})$ for the combination, we have $\bar{\alpha}(T' - T) = \Delta' - \Delta$ and $\delta\bar{\alpha}(T' - T) = \delta\Delta' - \delta\Delta$.

Evaluating these, we find from equations (1), (3), (4) and (5),

$$\bar{\alpha} = (\alpha_1 + \alpha_2)/2, \quad \dots\dots(6)$$

and (neglecting second powers of R in the numerator and first powers in the denominator)

$$\frac{\delta\bar{\alpha}}{\bar{\alpha}} = \frac{(\alpha_1 - \alpha_2)[(T_g - T')(R'_1 - R'_2) - (T_g - T)(R_1 - R_2)]}{2(\alpha_1 + \alpha_2)(T' - T)}. \quad \dots\dots(7)$$

It is also convenient to express this in another form, by multiplying out and refactorizing:

$$\frac{\delta\bar{\alpha}}{\bar{\alpha}} = \frac{(\alpha_1 - \alpha_2)[(T_g - T)(R'_1 - R'_2) - (T_g - T')(R_1 - R_2) - (T' - T)(R_1 - R_2 + R'_1 - R'_2)]}{2(\alpha_1 + \alpha_2)(T' - T)}. \quad \dots\dots(8)$$

We can now see under what conditions a thermal after-effect of the ordinary sign, that is, one for which $\delta\bar{\alpha}/\bar{\alpha} > 0$ is to be expected.

If we consider two glasses of differing composition, and take the case $T' > T$, then since R increases with temperature we have $R'_1 > R_1$ and $R'_2 > R_2$. Also if $\alpha_1 > \alpha_2$ it is usually the case that $R_1 > R_2$ and $R'_1 \gg R'_2$ (the sign $\gg$ is here used to denote a greater, though not necessarily very great, inequality). The correctness of the signs of these inequalities can be checked by considering pure silica as glass 2, when α_2 , R_2 and R'_2 would be very small.

In normal use, also, $T_g \gg T$ and $T_g \gg T'$. The sign of $\delta\bar{\alpha}/\bar{\alpha}$ is the same as the sign of the quantity inside square brackets in equations (7) and (8), and it can now be seen at once from equation (7) that this is positive.

By considering equation (8) we can see that there are in fact two effects acting in opposite directions. The first two products within the square brackets may be regarded as representing the fact that at the average stress acting between T and T' , the delayed elasticity of glass 1 preponderates increasingly as the temperature is raised. The difference between these products as written is always positive, and this is actually the main effect. The third product within the square brackets is also always positive and as it is prefaced by a minus sign, it represents a tendency towards a thermal after-effect in the reverse direction, which is due to the fact that when the combination is heated the stresses present are actually reduced. However, since $(T - T') \ll T_g$ this effect is relatively small; if the effect of increasing R_1 over-rides this reduction we have an after-effect of the usual sign.

§ 4. ESTIMATE OF THE DEPRESSION CONSTANT OF A MERCURY-IN-GLASS THERMOMETER

It may easily be shown that the depression constant k of a mercury-in-glass thermometer, as already defined, is given by

$$k = 300 \delta\bar{\alpha}/(\beta - 3\bar{\alpha}), \quad \dots\dots(9)$$

where β is the absolute coefficient of expansion of mercury, and $\bar{\alpha}$ and $\delta\bar{\alpha}$ now

refer to the material of the bulb between 0° and 100° . (There does not seem to be any objection to an extension of the foregoing argument to three dimensions.)

We have also $T = 0^\circ$, $T' = 100^\circ$, $T_g = 500^\circ$ approximately, for ordinary glasses. In order to substitute values of α and R we take average values as would be measured in the ordinary way for glasses of slightly differing compositions; the following would be reasonable values for, say, two soda-potash-lime-silica glasses of compositions differing by about 3–5% in the proportions of the main constituents: $\alpha_1 = 1 \times 10^{-5}$, $\alpha_2 = 0.8 \times 10^{-5}$, $R_1 = 0.01$, $R'_1 = 0.03$, $R_2 = 0.005$, $R'_2 = 0.01$. (Existing data do not permit of exact estimates, but the sense of the variation is correct, and the variation with temperature is of the right order). Also we have $\beta = 1.8 \times 10^{-4}$. Using the above numerical values in equations (6), (7), (9), we find $k \sim 0.05^\circ$. Choosing the following values to represent a case where the delayed elastic effect is only slightly greater for glass 1 and increases by only the same ratio at the higher temperature, also with a smaller difference between the coefficients of expansion: $\alpha_1 = 1 \times 10^{-5}$, $\alpha_2 = 0.9 \times 10^{-5}$, $R_1 = 0.01$, $R'_1 = 0.02$, $R_2 = 0.008$, $R'_2 = 0.016$, we find $k \sim 0.003^\circ$ and still of the correct sign. On the other hand, if we take values corresponding to pure silica for glass 2, and typical values for glass 1, the estimate of k becomes about 0.3° .

§ 5. EFFECT OF PRESSURE

It is worth considering whether the effects of changes in pressure on the volume of the bulb can play a part in the after-effects in mercury-in-glass thermometers, since they must always include delayed parts corresponding to the development of delayed elastic strains.

In ordinary use the most important change in pressure is that due to the increased height of the mercury column above the bulb when the thermometer is used vertically; this can amount to about 200 mm. of mercury at 100° in a thermometer calibrated from 0° to 100° . However, the "internal pressure coefficient" of such a thermometer is only about $0.0001^\circ/\text{mm. Hg}$ (Guillaume 1889), so that the effect of increasing pressure is only to lower the reading by about 0.02° at 100° . An additional delayed expansion of about 1–3% of this is to be expected, that is, corresponding to not more than about 0.0005° . It is easily seen that this would lead to "thermal after-effects" in the usual direction in the case of a thermometer used vertically, but of too small magnitude to be important. In gas-filled thermometers used to higher temperatures the effect would of course be greater. It is probably significant that reports have not been made of differences between the after-effects shown in the horizontal and vertical positions.

§ 6. CONCLUSION

We have thus accounted for the order of magnitude of the thermal after-effects by postulating a distribution of local composition corresponding to differences of only a few per cent in the relative proportions of the main constituents. Since large coefficients of expansion and large delayed elastic effects are usually found together—in glasses containing a high proportion of soda, potash, or both, for instance—it is to be expected that the effect will be in the direction usually noted. However, it is clear from equation (8) that effects in the reverse direction are not excluded by this explanation; they might also occur in glasses where the relationship between expansivities and delayed elastic effects is not as assumed here. It should be realized that the division of the material into two types of region which are present in equal abundance is no more than a mathematical

device, and the values of R must then be regarded as average values for a given type of region. If one assumed differing relative proportions of two types of region it would probably be justifiable to assume greater differences in the corresponding values of R so that the result would again be of the same order.

It seems probable that the effective structural variations are in fact inhomogeneities of composition. There is no need to justify this assumption of local variations in composition, which may perhaps be of colloidal dimensions; their existence can easily be demonstrated (see for instance Tooley and Tiede 1944), and repeated grinding and melting is actually necessary in the preparation of a glass required, say, for phase-equilibrium studies, where a higher degree of homogeneity is required than normally. It should be remembered, however, that variations in local structure are themselves ultimately responsible for the delayed elastic effects. A very clear picture of the thermal after-effects is possible if we attribute them to the presence of larger inhomogeneities than those which give rise to the delayed elastic effects, as assumed in the foregoing analysis. The general argument is, however, probably still valid if inhomogeneities of the same kind, possibly of very small dimensions, are actually responsible for both effects. These "inhomogeneities" might in fact simply be the irregular or amorphous atomic structure of the glass; it should be noted that the sizes of the regions do not appear in the analysis. It is not impossible that specially prepared homogeneous glass might show smaller after-effects of both kinds; as far as is known no experimental evidence yet exists on this point. (It should perhaps be stressed that we do not in any case consider the different regions as being separately composed of the individual chemical constituents—of silica and soda in a soda-silica glass for instance—but have considered the possibility of the mixture or solution of these constituents varying in composition from place to place.)

The relation described between the two after-effects offers a satisfactory explanation for the fact that a certain time is required at the higher temperature for the full zero-depression to be shown on subsequent cooling; in the recovery of delayed strains after removal of an ordinary mechanical stress only that part of the total is recovered which has had time to develop during the period of application of the stress. It also explains satisfactorily the fact that the effects increase with increasing temperature changes, though not always according to simple laws. There is a further point of interest, namely, that the zero-depression is much more persistent than, say, the "steam-point elevation", which usually disappears in about ten minutes and is generally not noticed. Similarly, the time required at the steam-point for full subsequent development of the zero-depression is also of this order. An explanation is possible if we realize that the glass cooled to 0° is actually compelled to recover delayed strains of magnitude greater than those normally corresponding to that temperature. It is not surprising that a long time is required, though as far as is known the direct mechanical counterpart of this behaviour has not been investigated in experiment. There is, however, a somewhat related phenomenon, known as "thermo-recovery", shown by some organic polymers, including polystyrene. After deformation this material recovers only partially, but on heating to a higher temperature further recovery occurs.

The reversible thermal after-effects might be expected to appear in polycrystalline metallic specimens in which crystallites of differing expansivities

occur together. However, the mercury-in-glass thermometer provides a very sensitive test for these changes, and this is probably the reason why they have been especially noted in glass. A somewhat related effect has of course been described for metals, known as "thermal fatigue", in which repeated changes of temperature can result in a permanent change in density.

The irreversible secular rise in the zero of a mercury-in-glass thermometer can be practically cured by heating the glass for a long period at a temperature rather lower than the normal annealing temperature, during which time the whole structure has an opportunity of approaching closer to its equilibrium condition at that temperature and the tendency to slow change at lower temperatures is thereby reduced or slowed down. This heating does not eliminate the reversible effects, however. The only ways in which it might be expected to reduce them would be by causing a slight reduction in the effective value of T_g , and by reducing the values of R on account of the closer approach to equilibrium conditions.

Finally, a word must be said about the somewhat analogous "thermal after-effect" shown by glass in the transformation region, where the equilibrium time of the process of stabilization is only of the order of minutes or hours. On rapidly raising or lowering the temperature of a specimen of glass maintained within this temperature region, there is an immediate change in volume, followed by a slow additional change in the same sense due to the slow configurational adjustments which are required in order to reach the degree of local order appropriate to the new temperature. (The values of $\delta\bar{\alpha}/\bar{\alpha}$ due to this cause in the transformation region would be much larger than those of the thermal after-effects at ordinary temperatures.) In the present paper we have considered the thermal after-effects only in relation to the delayed elastic effects, because their equilibrium times are of the same order, and short, while the equilibrium time of the stabilization process at ordinary temperatures is probably at least of the order of hundreds of years. However, the distinction is not fundamental, and in the transformation region, where the equilibrium times of the stabilization process and the delayed elastic effects are of the same order, the two phenomena can probably be quite closely related, as suggested by Frenkel (1946). All the effects which we have considered are macroscopic manifestations of atomic migratory processes analogous to those occurring during ordinary viscous flow. In the absence of a more complete quantitative picture, the foregoing analysis is sufficient to explain how a material which shows a delayed elastic effect may also show a thermal after-effect.

ACKNOWLEDGMENT

The author wishes to acknowledge the helpful comments of Dr. E. Orowan, who read this paper before it was submitted for publication.

REFERENCES

- CRAFTS, J. M., 1882, *C.R. Acad. Sci. Paris*, **94**, 1298.
- DICKINSON, H. C., 1906, *Bull. Bur. Stand. Wash.*, **2**, 189.
- FRENKEL, J., 1946, *Kinetic Theory of Liquids* (Oxford: University Press).
- GUILLAUME, C. E., 1889, *Traité pratique de la thermométrie de précision*.
- HOVESTADT, H., 1902, *Jena Glass* (New York: Macmillan).
- JONES, G. O., 1944, *J. Soc. Glass Tech.*, **28**, 432.
- TOOLEY, F. V., and TIEDE, R. L., 1944, *J. Amer. Ceram. Soc.*, **27**, 42.
- WEIDMANN, G., 1886, *Ann. Chim. (Phys.)*, **29**, 214.
- YOUNG, S., 1890, *Nature, Lond.*, **41**, 488.

The Initiation of Low Pressure Glow Discharges

By J. D. CRAGGS* AND J. M. MEEK†

* Research Department, Metropolitan-Vickers Electrical Co. Ltd.

† Department of Electrical Engineering (Electronics), University of Liverpool

MS. received 14 January 1948

ABSTRACT. Oscillographic studies of some transient low pressure gaseous discharges have been made, with particular attention to the intermittent discharges occurring before stable self-maintained glow discharges are established. The nature of these pre-glow discharges varies considerably with the nature of the gas, and attempts are made to interpret them in relation to Weissler's corona work.

§ 1. INTRODUCTION

NUMEROUS studies have been made by various investigators of the breakdown voltage/pressure relation in gases at low pressures. Huxley, Penning, Bruce and Boulind, among others, have studied He, Ne, N₂, A, H₂, O₂, CO₂ and air. In their measurements the criterion for breakdown has been the sharp increase in current, as measured by a galvanometer, which occurs when breakdown commences. The increase is so abrupt (Druyvesteyn and Penning 1940) that the onset voltage is clearly defined within a narrow range, usually smaller than the accuracy of measurement. None of the above investigators reported the current-time fluctuations in the low-pressure discharge; this appears to have been first studied by Bennett (1914). Sloane (1937) investigated the discharge characteristics oscillographically for point/point breakdown in hydrogen. Extensive studies at higher pressures (~1 atmosphere) have been made by Loeb and his collaborators with oscillographs.

In the present experiments the onset of glow discharges in various gases at pressures less than 30 mm. Hg has been investigated and the results are presented in the form of oscillograms. The use of A.C. (50 c/s.) is unusual in work of this kind, though some qualitative measurements have been made in similar ways by Bennett (1914) and Tykociner *et al.* (1935), mostly for higher pressures.

The use of A.C. has some important advantages in the oscillographic measurements, but care must be taken in the interpretation of the results to distinguish between effects caused by the discharge characteristics and those which might be attributed to the alternating character of the voltage. The A.C. results have therefore been frequently checked with D.C. and, with the exception of wire/cylinder discharges in argon, it is unlikely that any but very minor effects are due to the use of A.C.

The D.C. breakdown measurements of other authors have been repeated with close agreement, as a further check on the experimental methods.

§ 2. BREAKDOWN VOLTAGES

The circuit used for the determination of breakdown voltage, and also for the later observations of the discharge characteristics, is shown in figure 1. The

breakdown voltages (figures 3-7) were measured with an accurate electrostatic voltmeter V and a potential divider R_2 , and are considered accurate to $\pm 2\%$. After breakdown, these measurements do not give an accurate record of the gap voltage, and this was obtained oscillographically, as shown in figure 1. The

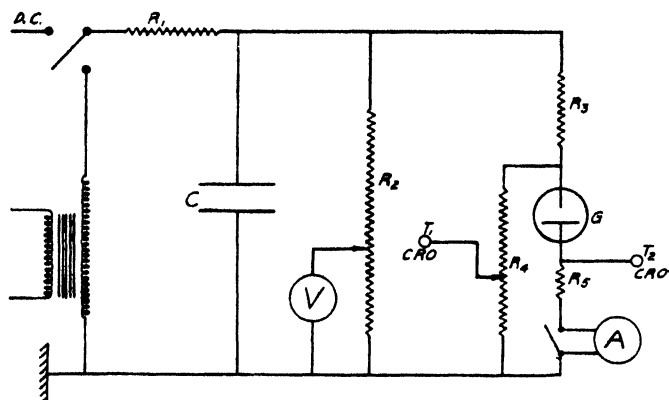


Figure 1. Circuit used for observation of onset of glow discharge at low pressures.

presence of the capacitance C ensured that the total A.C. voltage across R_3 and the discharge tube was not affected by the discharge current. Currents were measured oscillographically by observation of the voltage developed across R_5 . In the D.C. measurements the average discharge current was measured with the microammeter A .

The explanation of the results is facilitated by reference to the general diagram (figure 2) of voltage/current relations in a discharge (Penning and Druyvesteyn 1940). As the voltage is raised breakdown occurs at B with a sudden increase in current up to 10^{-6} amp. or more; V_B is referred to as the breakdown voltage for the discharge. With further increase of current there is a region of transition

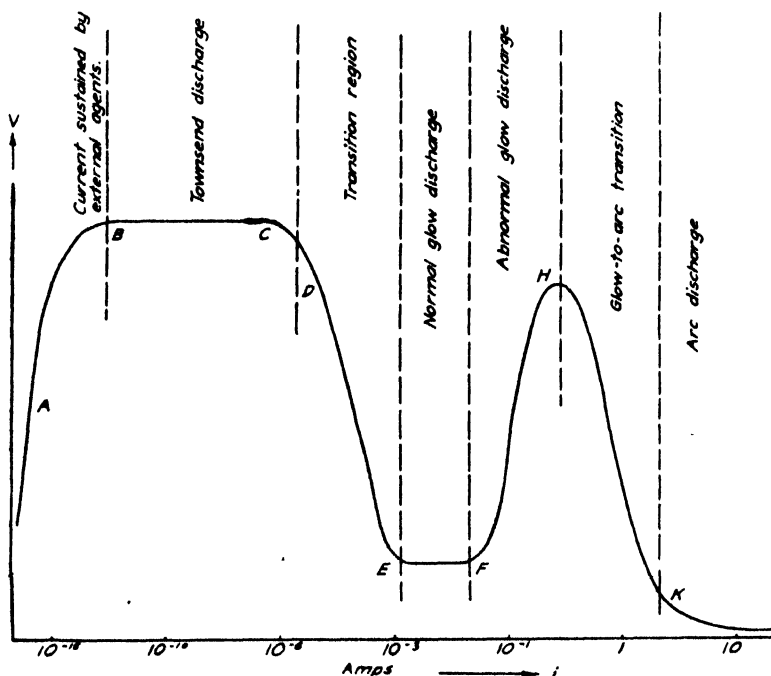


Figure 2. Representative characteristic curve of gaseous discharge.

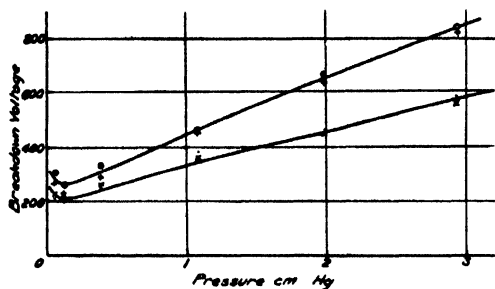


Figure 3. Breakdown voltage/pressure for argon (1/16 in. diameter round point/plane.)

- D.C. +ve.
- × D.C. -ve.
- + A.C. +ve.
- A.C. -ve.

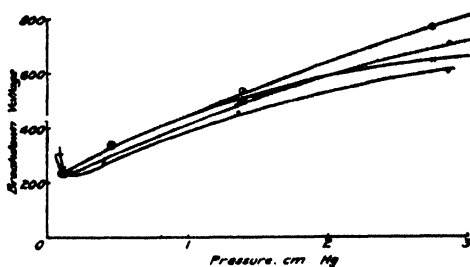


Figure 4. Breakdown voltage/pressure for argon (wire/cylinder).

- D.C. +ve.
- × D.C. -ve.
- + A.C. +ve.
- A.C. -ve.

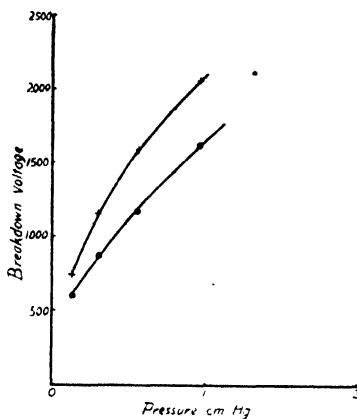


Figure 5. Breakdown voltage/pressure for CCl_2F_2 (ficon) (wire/cylinder).

- + A.C. +ve.
- A.C. -ve.

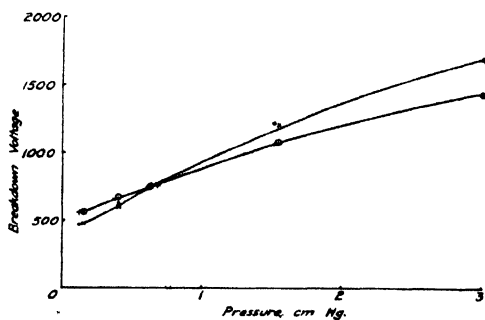


Figure 6. Breakdown voltage/pressure for oxygen (wire/cylinder).

- D.C. +ve.
- × D.C. -ve.
- + A.C. +ve.
- A.C. -ve.

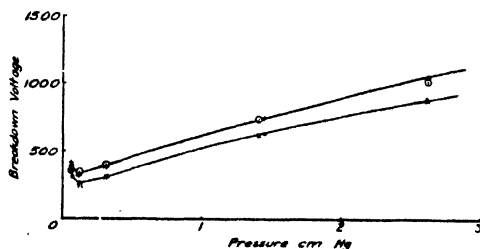


Figure 7. Breakdown voltage/pressure for hydrogen (wire/cylinder).

- D.C. +ve.
- × D.C. -ve.
- + A.C. +ve.
- A.C. -ve.

DE and a region EF of fully established glow discharge. In the present experiments the region CDE is of major interest.

Various forms of discharge gaps (point/plane and wire/cylinder) were studied, but in most of the experiments the electrodes consisted of a tungsten wire 0.126 cm. in diameter surrounded by a coaxial copper cylinder of 2.5 cm. internal diameter mounted in a glass tube. The discharge space was about 12 cm. in length. The electrodes were chemically clean, but no precautions were taken to free the apparatus from impurities by continued evacuation at high temperatures. The composition of the gases used was such that, with the exception of methane which was not analysed, a purity of at least 99.5% was obtained. Some of the argon included 2% of nitrogen, but spectroscopically pure argon behaved similarly. An Apiezon oil manometer was used to avoid mercury vapour contamination.

Curves showing the variation of breakdown voltage with pressure for different gases (A.C. and D.C.) are given in figures 3–7. The point/plane gap of figure 3 was 1 cm. in length and the point was ground roughly spherical and of radius approximately 1 mm. In all the gases used, with the exception of argon, there is no detectable difference between the positive D.C. breakdown voltage and the A.C. peak voltage which causes discharge onset on the positive half-cycle. Similarly no difference is observed between the negative D.C. and A.C. breakdown voltages. This indicates that the time between the crests of successive half-cycles is sufficient to allow the clearance of ions from the gas and so to prevent the influence of space charge created on one half-cycle from affecting the subsequent half-cycle. In argon, however, the A.C. breakdown voltage is appreciably lower than the D.C. value, when the wire/cylinder arrangement is used.

It was also observed in argon that small currents flow during the positive half-cycles if appreciable currents are established during the preceding negative half-cycles. These effects, we suggest, may be attributed to the persistence of metastable atoms in argon, but no similar effects were detected in the other gases used, some of which (N_2 , He and Ne) are known also to possess metastable states. In these latter cases it would therefore appear that the lifetimes of the metastable atoms are short although the selective ionization of suitable impurities (Penning effect) may also have been greater in argon than in the other gases mentioned.

The values of the breakdown voltages measured agree closely with those of other workers. Direct comparison of the V - p curves is not possible because of the different electrode diameters used, but it is convenient to compare curves relating rX and rp where r is the wire radius and X is the voltage gradient at breakdown at the wire surface. In the latter curves, provided the diameter of the outer cylinder is large compared with the wire diameter, measurements of breakdown voltage as a function of pressure give values lying on the same curve for any given gas. The validity of this method of comparison of breakdown gradients is discussed by Huxley (1928).

Figures 8–11 show curves of this type and include similar curves obtained by other investigators (Penning 1931, Huxley 1928, 1930, Boulind 1934, 1935). In argon (figure 8) the curve for a positive wire is in close agreement with that obtained by Penning, but there is an appreciable difference between the two curves for a negative wire, probably on account of the difference in wire materials used. The latter difference will not affect the breakdown for a positive

wire (Loeb 1939, p. 511) for which the criterion for discharge onset is governed, at these pressures, by photo-ionization in the gas or by electron emission at the cylindrical cathode, and is given by

$$\eta\theta g/\bar{\alpha} \exp\left(\int_r^R \alpha dx\right) = 1. \quad \dots\dots(1)$$

Here α is the ionization coefficient for collision by electrons in the gas, $\bar{\alpha}$ is the mean value of α near the wire, θ is the number of photons produced by an electron

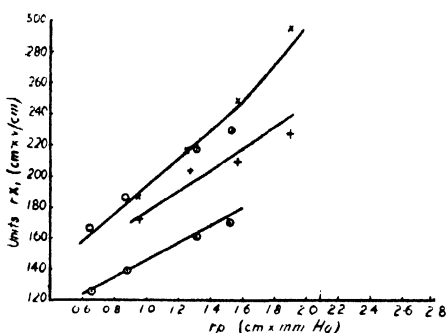


Figure 8. Breakdown characteristics for argon (wire/cylinder).

- +ve {
○ Penning Ni wire.
× present work W wire.
-ve {
○ Penning.
+ present work.

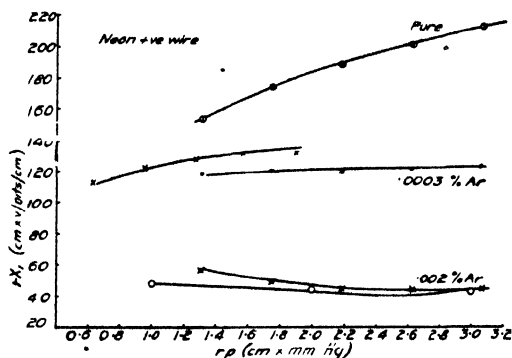


Figure 9. Breakdown characteristics for neon (+ve wire/cylinder).

- {
● Penning's results.
× Present work.
○ Huxley.

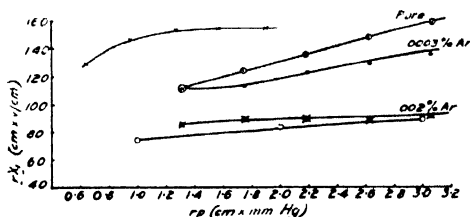


Figure 10. Breakdown characteristics for neon (-ve wire/cylinder).

- {
○ Penning's results.
● Present work.
○ Huxley.

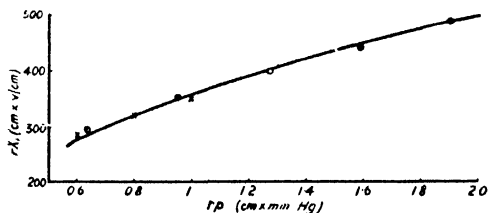


Figure 11. Breakdown characteristics for oxygen (+ve wire/cylinder).

- {
○ Present work.
× Boulind's results.

per cm. of path in the direction of the field, η is the fraction of photons which produce electrons that succeeds in leaving the cathode, and g is a geometrical factor. The wire material has therefore no effect on the onset of the positive discharge, but will have considerable influence on the negative wire breakdown because of the resultant change in the secondary ionization coefficient which largely governs the onset of the negative discharge. The criterion for onset at the negative wire is generally taken as

$$\exp\left(\int_r^R \alpha dx\right) = 1 + (1/\gamma). \quad \dots\dots(2)$$

γ is the number of secondary electrons leaving the cathode per incident particle or quantum.

It is clear that the lower the value of γ for the wire cathode the higher will be the voltage gradient required to cause breakdown of a given gas at a given pressure. Such considerations probably explain the difference between the curve obtained by Penning, who used a nickel wire, and that of the present authors, whose wire was of tungsten. No measurements comparing the value of γ for these materials in argon appear to be given in the literature.

Similar considerations apply to the other gases, where there is also closer agreement between the positive breakdown values than between those for negative breakdown. Additional interest is attached to the neon results because of Penning's experiments on the effects of contamination of this gas with small amounts of argon. As shown in figure 9, which gives some of the curves obtained by Penning, an amount of only 0.0003% of argon in the neon alters appreciably the breakdown voltage curve. The effect is attributed to the presence of metastable neon atoms, whose excitation potential is in excess of the ionization potential of argon, and which consequently exert an important influence on the breakdown voltage for the gas. Penning's results indicate that Huxley's measurements must have been made in neon which probably contained argon. In the present experiments it would appear that the neon used was comparatively pure, with less than 0.0003% argon, as shown by comparison of the curves for the positive wire. The results for the negative wire are again difficult to compare, as explained above, on account of the use of tungsten rather than nickel wire, but tend to confirm the purity of the gas and the lower value of γ for tungsten.

Figure 11 shows the extremely close agreement between Boulind's results and those of the present experiments for positive discharges in oxygen. The hydrogen results also show the same satisfactory agreement. In these tests minor impurities are apparently not important.

In agreement with the results of other workers (Loeb 1939, p. 513) breakdown at the reduced pressures used in these experiments generally occurs at a lower voltage when the smaller electrode is negative than when it is positive. This is true for all the gases used except oxygen (see figure 6) and air, as Boulind (1934) and others had previously shown. The reason for these exceptions is not altogether clear, but it is probably not entirely due to the high electron attachment coefficient in oxygen as suggested by previous authors because, in the case of partially dissociated freon,* where the attachment coefficient is also high because of the presence of free halogens, the negative breakdown curve lies below the positive curve (figure 5). It is possible, however, that the work function of the cathode may be appreciably affected in the presence of oxygen, either by an absorbed oxygen layer or by oxidation, and calculations based on these experiments indicate that the value of γ at onset of the negative breakdown is much lower in oxygen than for the other gases. In these calculations the usual criterion $\exp\left(\int_r^R \alpha dx\right) = 1$ † is assumed to define the negative breakdown, and the value of $\int_r^R \alpha dx$ has been determined for the various gases in the negative breakdown field at 30 mm. Hg pressure. The resultant calculated values of γ are given in table 1, using breakdown voltages from figures 3–11. The values of $\alpha/p(f(X/p))$ for hydrogen, argon, nitrogen and oxygen respectively, which were used in the calculations of γ given in table 1, were taken from the papers of Hale

* CCl_2F_2 , known commercially as a refrigerant, "Freon 12".

† In practice, see table 1, $\gamma \ll 1$.

(1938, 1939), Kruithof and Penning (1936, 1937), Bowls (1938), and Posin (1936) for nitrogen, and Masch (1932).

Table 1

Gas	X/p (v/cm/mm.)	γ	γ (Loeb)
Hydrogen	173	25×10^{-3}	25×10^{-3}
Argon	120	4.5×10^{-3}	15×10^{-3}
Nitrogen	285	2.5×10^{-3}	20×10^{-3}
Oxygen	295	2×10^{-5}	—

The breakdown voltage for nitrogen at 26 mm. Hg pressure was 1400.

The last column gives results from Loeb (1939, pp. 392 *et seq.*) for γ with a Pt surface. The above values of γ may be considered as correct in order of magnitude and indicate a very much lower value of γ in oxygen than in the other gases. The relevance of this result does not seem to have been appreciated in earlier work. Some confirmation of it is obtained by the authors' calculations based on the results of Huxley (1928, 1930) and Boulind (1934, 1935) who found that the negative breakdown potentials in N_2 and O_2 at 6 mm. Hg for a nickel wire (1.65 mm. in diameter) in a cylinder 46 mm. in diameter were 515 and 960 volts respectively. The calculated values of γ are 0.18 and 2.8×10^{-4} . A full discussion of the significance of γ and its interpretation when used in the conventional breakdown formula (given above) for low pressure gaseous conduction is given by Druyvesteyn and Penning (1940) and Loeb (1939).*

§3. NATURE OF CORONA EFFECTS

The main purpose of the present experiments was to examine the discontinuities occurring in the region CDE of figure 2. The oscillographic technique was not suitable for investigation of the Townsend discharge, region BC of figure 2, which may also be expected to exhibit discontinuities (electron avalanches). Currents in the region BC are $1 \mu a.$ or less, whereas currents in the discontinuous region of the glow discharge onset may be 10 – $100 \mu a.$

Earlier investigators, e.g. Bruce (1930), Boulind (1934) and Penning (1931), used galvanometers of sensitivity approximately 6×10^{-11} amp/mm. to detect breakdown and did not observe the current wave-forms. Sloane (1937) obtained results in hydrogen which showed the presence of certain regions in the i/v curves where static characteristics could not be obtained, and oscillographic measurements were made.

In this section is described the nature of the current variations observed with the circuit of figure 1 when the voltage is above that defined as the breakdown voltage in §2. Current values are 5 – $500 \mu a.$

The types of current variation may be briefly summarized. (a) The current varies smoothly with voltage from 1 – $5 \mu a.$ (not measurable) to a value occurring at the peak of the applied A.C. voltage wave: see for example the negative current of figure 16(c), and the positive current of figure 15(c). With D.C. the same

* The authors are greatly indebted to Professor L. B. Loeb for pointing out that the calculation of γ in high fields may sometimes be uncertain unless Morton's (1946) work on α in fields where X varies appreciably over one electron mean free path (L_e) is considered. Morton shows that if $\phi = (L_e/X)(dX/dx) < 2.5\%$, then the error introduced when the Townsend α is used $< 25\%$. For the field at the wire in the present arrangement ϕ varies from about 1% to 4% for the various gases, so that the error introduced in the integrated values of α from wire to cylinder used for the calculation of γ (see equation (2)) will not be significant. The γ values in table 1 should only be taken as accurate to one significant figure.

effect is observed. (b) The current records often show time variations of the relaxation oscillation type, similar in general appearance to the Trichel pulses (Trichel 1938, 1939) observed in high pressure negative corona and to those observed with the flashing neon lamp (resistance/capacitance circuit). Figure 15 (d) shows a D.C. record with a ripple due to the rectifier circuit; with A.C. it is sometimes observed, as in figure 17 (positive), that with increased voltages the extinction current is higher than when lower voltages are used. Sometimes (figure 21 (c), positive) several modes of this type of current variation occur. Current oscillations may also be observed at low currents, whilst at higher currents for the same

Table 2. Details of oscillograms shown in figures 12-15

Figure	Gas	Pressure (cm. Hg)	V_p (v.)	V_c (v.)	I (ma/cm.)	R_s ($\Omega \times 10^5$)	t (sec.)
12 (a)	Argon	2.84	1020	+ 700 - 620	0.70	1.02	1/100
12 (b)	"	1.37	520	+ 485 - 445	"	"	1/25
12 (c)	"	"	570		"	"	"
12 (d)	"	"	780		"	"	"
13 (a)	Argon	0.40	337	+ 285 - 270	0.098	7.3	1/25
13 (b)	"	"	457		"	"	"
13 (c)	"	"	521		0.70	1.02	"
13 (d)	"	"	594		"	"	1/100
14 (a)	Argon	0.13	315	+ 240 - 240	0.052	1.38	1/25
14 (b)	"	"	322		"	"	"
14 (c)	"	"	343		"	"	"
15 (a)	Hydrogen	2.66	1140	+ 1060 - 890	0.70	1.02	1/100
15 (b)	"	"	1200		"	"	"
15 (c)	"	"	1530		"	"	"
15 (d)	"	"	1097		"	"	D.C.
15 (e)	"	"	1138		"	"	D.C.

V_p =peak voltage applied to (R_s +tube) ; V_c =corona onset (or breakdown) voltage ;
 I =current sensitivity on records ; t =total c.r.o. sweep time.

Positive currents are above the centre line on the oscillograms. Reference to figures 22 and 23 will aid in an understanding of figures 12-15. The latter are 3/5 full size.

records they are not seen, even when the current falls towards zero after passing the peak. (c) A discontinuity in a current wave of type (a) or (b) is observed, after which a smooth current variation is observed, see figure 19 (c) (positive). Type (c) may be safely interpreted as showing a transition to a stable glow discharge. It is likely that case (b) often corresponds to a relaxation oscillation caused by a falling discharge characteristic and a negative total circuit resistance (Schulze 1932 and references there cited). The current discontinuities may be due to space charges the result of which is to choke off the discharge, which subsequently re-starts, as the work of Kip and Trichel has shown. It is not easy to separate these effects with certainty because, while circuit constants may have an effect on the frequency of relaxation oscillations caused by the

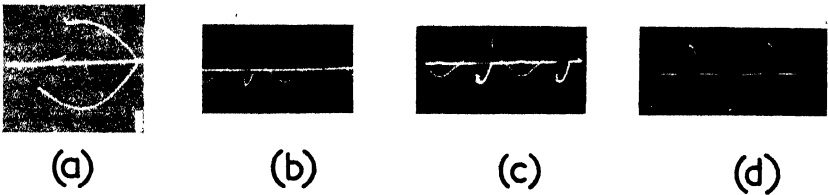


Fig.12.

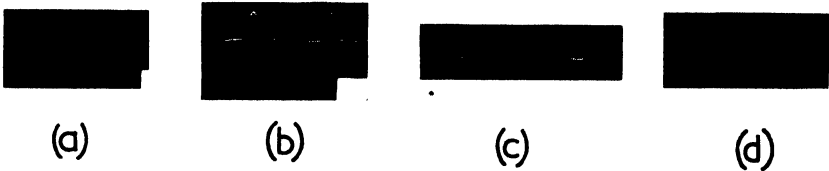


Fig. 13.

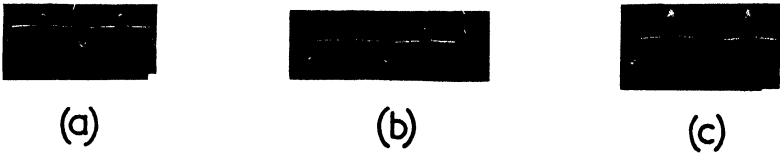


Fig. 14.

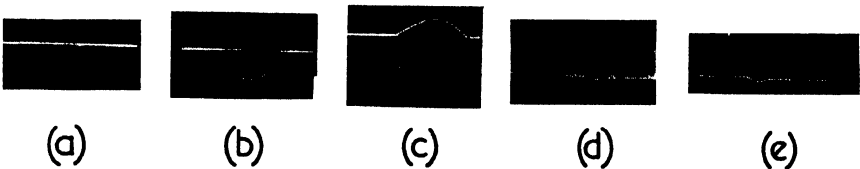


Fig. 15.

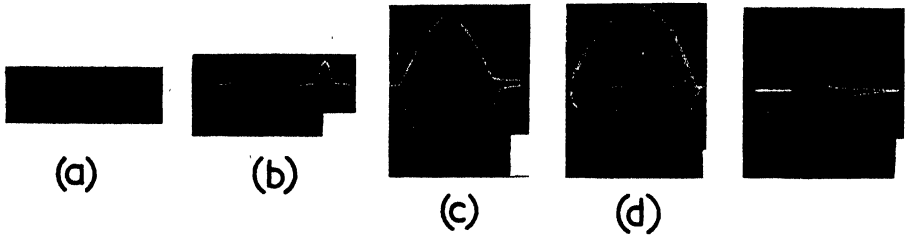


Fig. 16.

Fig. 17.

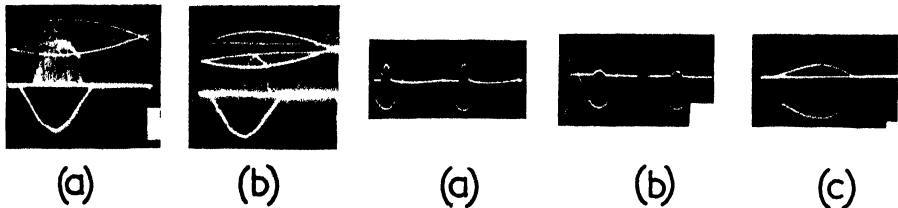


Fig. 18.

Fig. 19.



Fig. 20.

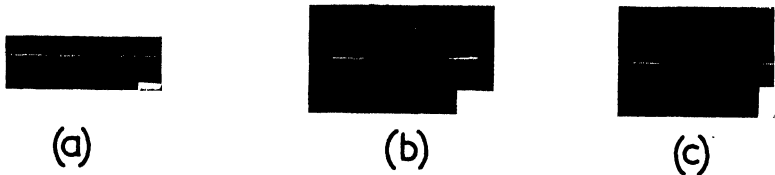


Fig. 21.

difference between striking and extinction voltages and the existence of a negative circuit resistance, these same constants may also affect the space charge phenomena. It is hoped in later work with much longer gaps and correspondingly low circuit resistances to distinguish clearly between these effects, and to assess the effect of circuit parameters in general on the observed nature of the oscillations.

It is difficult to account for case (a). The finite resolving power of the oscillograph may not allow rapidly varying currents to be accurately observed, so that only a mean current is measured. It is possible in some cases that a Townsend current is being observed (see for instance Druyvesteyn and Penning

Table 3. Details of oscillograms shown in figures 16-21

Figure	Gas	Pressure (cm. Hg)	V_p (v.)	V_c (v.)	I (ma/cm.)	R_s ($\Omega \times 10^6$)	t (sec.)
16 (a)	Hydrogen	0.14	382	+ 340 - 260	0.098	7.3	1/25
16 (b)	"	"	485		"	"	"
16 (c)	"	"	688		"	"	1/100
16 (d)	"	"	1732		0.70	1.02	"
17	Methane	0.20	770	+ 960 - 678	0.098	7.3	1/100
18 (a)*	Freon	0.70	1570	+ 1760	0.22	3.28	1/100
18 (b)†	"	"	"	- 1350	"	"	"
19 (a)	Oxygen	1.54	1340	+ 1070 - 1170	0.098	7.3	1/25
19 (b)	"	"	1370		0.70	1.02	"
19 (c)	"	"	1800		0.15	4.8	1/100
20 (a)	Oxygen	0.68	870	+ 770 - 770	0.098	7.3	1/25
20 (b)	"	"	960		0.70	1.02	1/100
20 (c)	"	"	1158		"	"	"
20 (d)	"	"	2020		0.15	4.8	"
21 (a)	Oxygen	0.38	744	+ 660 - 600	0.70	1.02	1/100
21 (b)	"	"	1057		"	"	"
21 (c)	"	"	1200		"	"	"

V_p =peak voltage applied to (R_s +tube) ; V_c =corona onset (or breakdown) voltage ;
 I =current sensitivity on records ; t =total c.r.o. sweep time.

Positive currents are above the centre line on the oscillogram.

* Without radium } Note that measurements of the positive corona onset voltage are
 † With radium } unsatisfactory owing to the strong effect of the radium.

1940, p. 104), but the magnitudes of currents of type (a) are often too great, as in figure 16(c), for this explanation to be tenable.

The oscillographic results for wire/cylinder gaps having the electrode sizes given in § 2 show the following characteristics: (a) with argon (figures 12-14) only disruptive discharges are noticed at the higher pressures (10-30 mm. Hg), as in figure 12(a). The positive discharges often start smoothly. At the lower pressures, oscillatory discharges occur (figure 14(c)) usually superimposed on the smooth discharges. Neon and helium tend to favour the production of oscillatory discharges in preference to the disruptive or continuous types.

(b) Hydrogen (figures 15 and 16) gives oscillatory currents at all but the lowest pressures, i.e. below the Paschen minimum. At 30 mm. Hg and with a sharp point, smooth discharges are noticed at low currents. The correlation between A.C. and D.C. records is shown by figures 15(d) and 15(e). Some of the records also show oscillograms of tube voltage, e.g. figure 15(b). (c) Oxygen (figures 19–21) gave oscillations at all but the lowest pressures, but the negative discharges were smooth at the higher pressures. The unusual effect shown in figure 23(c) was most prominent; there appear to be three modes of discharge, of a similar nature, which occur one by one as the current is increased. Discharges in air also show these features. Figure 22 shows an enlargement of figure 19(b) taken with oxygen. The oxygen results also show, figure 19(a), that positive

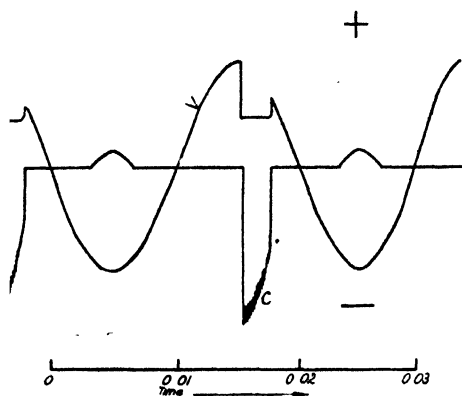


Figure 22. Current/time and voltage/time A.C. oscillogram for oxygen (wire/cylinder). Enlargement of figure 19(b), see table 3 for data.

V, voltage curve; C, current curve.

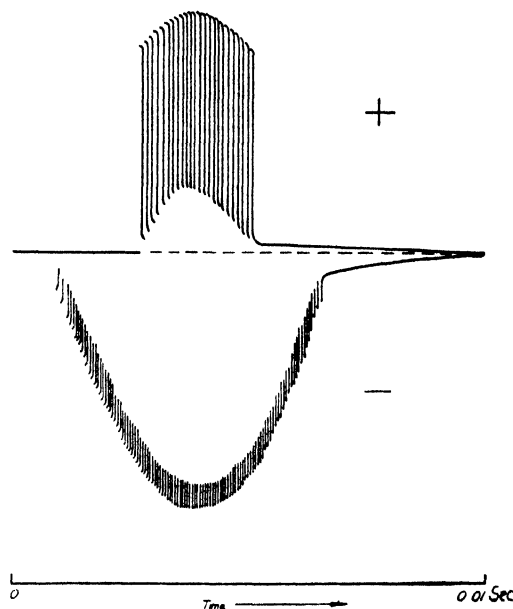


Figure 23. Current/time A.C. oscillogram for methane (wire/cylinder). Enlargement of figure 17, see table 3 for data.

breakdown may occur at lower voltages than negative breakdown, which is a characteristic property of oxygen (figure 6). Disruptive discharges are noticed in oxygen at the higher pressures (figure 19(c)). Boulind (1934) detected disruptive discharges by observation, not with an oscillograph but by the less satisfactory method of detecting the large voltage drop introduced into the circuit by such discharges. It was stated (Boulind 1934) that negative discharges in oxygen were always disruptive, but the present investigation does not confirm this conclusion for pressures below about 2 mm. Hg (figure 21(a)). (d) Nitrogen gave oscillations at all pressures observed. The wire cylinder gap gave smooth positive and negative discharges only at the highest currents and pressures. (e) Freon (CCl_2F_2) (figure 18) was studied because of the general interest in this gas as an insulator. Disruptive or continuous discharges were prominent and the oscillatory types were usually observed only at the lower pressures. The great effect of irradiation with radium gamma-rays is shown in figures (18 a)

and 18(b), taken respectively with and without irradiation. (In the other gases, for the conditions of current etc. obtaining in these experiments, such irradiation had negligible effect on the records. However, 0.2 mg. of radium was placed a few millimetres from the tube for all the experiments described in this paper). There would appear to be such a shortage of free electrons, presumably because of the high attachment coefficient in dissociated freon, that gamma-rays give rise to observable effects even in the presence of intermittent positive discharges (figures 18(a) and 18(b)) of ~ 0.2 ma. and stable negative discharges also of ~ 0.2 ma. It is to be expected that irradiation would affect the positive discharges because of the relatively large cathode area (the cylinder), while for the negative wire discharges the field at the wire will be high enough to prevent attachment from being effective. (f) Methane (figure 17) was studied as an example of a complex molecular gas. Oscillatory discharges were noticed, except at the highest pressures, when continuous positive discharges, becoming disruptive, and disruptive negative discharges were observed. Figure 17 is shown enlarged as a particularly good example of the oscillatory type of discharge (figure 23).

Experiments, using several of the above gases, were made with D.C. Invariably the same sequence of events was observed, showing that the use of A.C. had no important effect on the nature of the phenomena observed, as discussed in previous communications (Craggs and Meek 1943 a and 1943 b). The lowered A.C. breakdown with argon has been discussed in §2.

As the use of A.C. renders the presentation of results easier, the use of D.C. was discontinued except for purposes of checking the A.C. results.

§4. DISCUSSION OF RESULTS

Figure 24 shows drawings, not to scale, of some of the D.C. discharges. A point/plane gap was used, for convenience in observation, as there seemed to be no significant differences in the character of discharges in wire/cylinder and point/plane gaps. It seems clear from figure 24(i) that the disruptive discharges are normal glows, but that the oscillatory discharges show in general a fairly uniform illumination of the whole gap. Both kinds of discharge often occur together. The spinning discharges are probably caused by time variations of γ at the cathode surface due to bombardment. It seems that none of the discharges observed in these experiments were of Townsend type since the currents would then be less than $1 \mu\text{a}$. Such discharges, if observable visually, would presumably show continuous illumination in the gap.

Valle (1926) and others have pointed out that intermittent discharges may be expected for the falling part (CDE) of figure 2 if the total circuit resistance is negative (Cobine 1931, pp. 207–209), i.e. if

$$de/di + R < 0, \quad \dots\dots(3)$$

where de/di is the slope of the discharge characteristic at the point of its intersection with the voltage/current line for the circuit resistance R . This is not in itself an explanation of the oscillatory discharges, since it is necessary to show how such a negative characteristic is produced and to explain its variation with different gases in terms of the movements of ions and electrons. Schulze's (1932) description of the discharge path as having a natural impedance, i.e. reactance and resistance, seems inadequate.

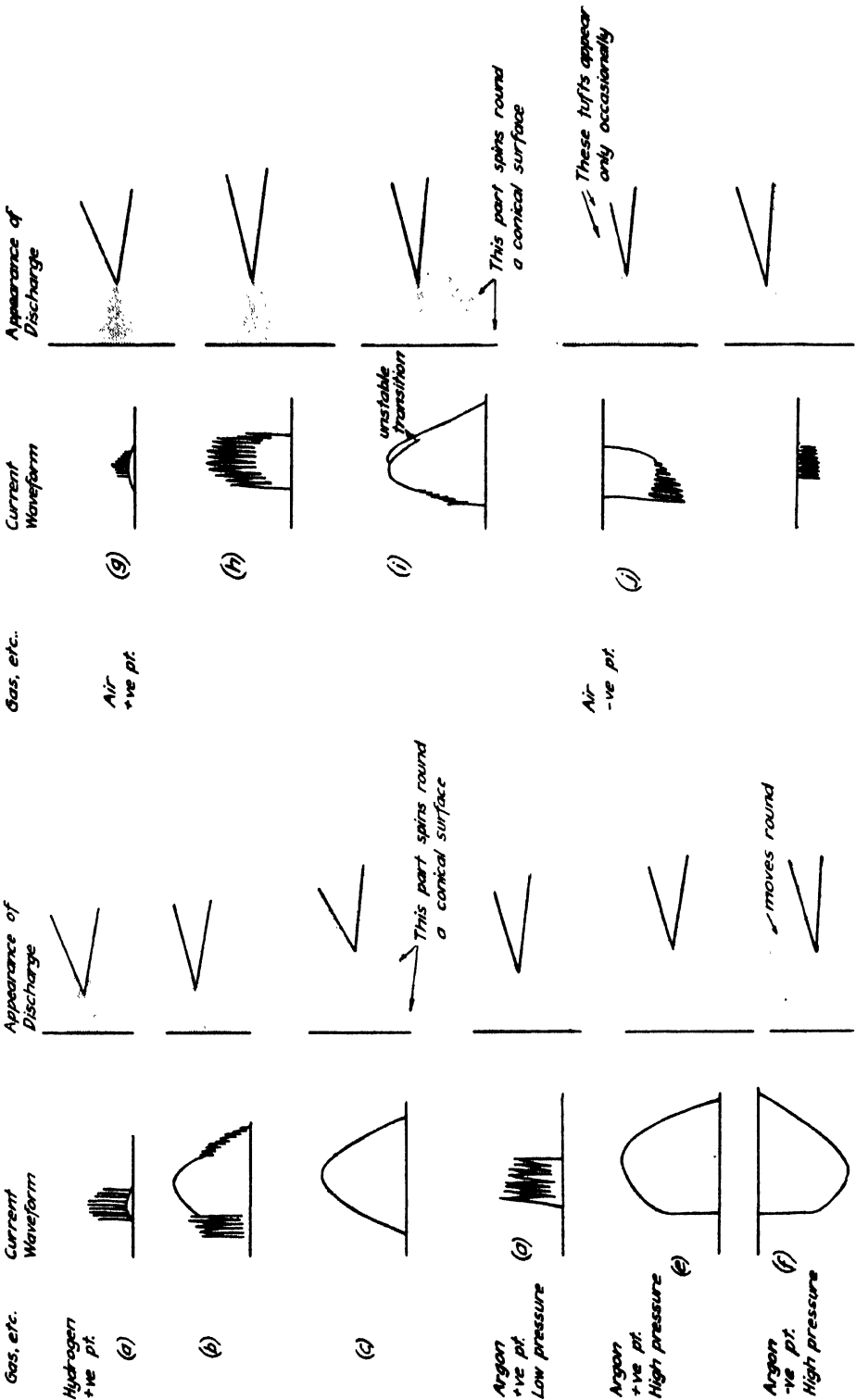


Figure 24. Appearance of discharges in argon and hydrogen for different types of current wave-form.

Penning (1931) and Penning, Moubis and Addink (1933) have studied argon discharges in a wire/cylinder gap in which the radii of wire and cylinder were respectively 0.087 cm. (nickel) and 2.3 cm. (iron). For a positive wire the effect of space charges became apparent, when the pressure was 12.7–197 mm. Hg, at 1–10 μ a. and the characteristic became positive. For lower pressures the characteristic was negative. (In the Townsend régime space charge effects are not encountered.) These and other effects are explained as follows (Druyvesteyn and Penning 1940).

For positive corona, an increasing current leads to accumulation of positive space charge round the wire where the field is thus weakened. So long as the field X at the cathode is too low for the Townsend coefficient α to be appreciable, it follows that an increase in the voltage V is necessary for maintenance of the increased current. If X becomes appreciable at the cathode, the space charge augments the total ionization in the gap and V passes through a maximum, and the characteristic becomes negative so that a glow discharge is obtained.

Continuing the arguments of Druyvesteyn and Penning (1940, p. 124), it follows that in electron-attaching gases such as oxygen, air, Freon etc. at fairly high pressures, so that X/p at the cylinder is too low to ensure detachment, the negative corona has a positive characteristic. Corona may for the present purpose be defined as a discharge lying between the Townsend and glow discharges. For non-attaching gases, e.g. argon, negative corona has a negative characteristic at the higher pressures, from the above discussion. Penning's results for argon, quoted above for positive corona, showed that the negative corona had a negative characteristic for pressures greater than 1 mm. Hg with transitions to intermittent discharges, and a positive characteristic for pressures less than 1 mm. Hg.

These conclusions for argon are supported by the results given above (figures 12–14) which show immediate transitions to a glow discharge, i.e. disruptive discharges, for a negative wire at higher pressures, while the positive discharge is stable up to large currents as shown in figures 12(a) and 12(c).

The positive corona in hydrogen is continuous, at the higher pressures, to greater currents than in argon, in agreement with Weissler's (1943) conclusion that complete breakdown to a spark at pressures ~ 1 atmosphere (or, as in the present case, to a glow at low pressures) took place much more readily in argon, where the presence of metastable atoms leads to more efficient photo-ionization than in relatively photon-transparent gases such as hydrogen.

Negative discharges in hydrogen are also seen to be more continuous than in argon, for the same general reasons as those given above, although the discontinuous type of negative discharge in hydrogen replaces the disruptive negative discharges in argon. Photo-ionization, discussed by Weissler, is more important at higher pressures, say about 1 atmosphere; at lower pressures, such as those used in the present work, mechanisms at the cathode are more important. Weissler found that positive and negative breakdown occurred at approximately the same voltage in argon, but not in hydrogen, at 1 atmosphere, since electron generation in the gas is the controlling process for both negative and positive discharges at these pressures. The present results, taken at much lower pressures than Weissler used, do not show this effect (see figures 4 and 7).

The oxygen discharges do not show, except at the lowest pressures, the continuous positive discharge shown at high pressures (~ 3 cm. Hg) by hydrogen or argon. This is possibly because the formation of negative ions by attachment.

does not allow positive space charges to build up round the wire and so produce the positive characteristic in the manner explained above.

§ 5. CONCLUSIONS

The use of low frequency A.C. for the oscillographic observation of low pressure breakdown effects in gases has been shown to lead to simplifications of technique.

The method described has also shown that such breakdown processes differ considerably with the nature of the gas investigated. In certain gases, e.g. argon, breakdown usually leads to the formation of a fully established glow discharge without intermediate intermittent discharges. In other gases, e.g. hydrogen, a great variety of intermittent discharges are observed. Some of these are qualitatively explicable and their connection with negative discharge characteristics is indicated. Photo-ionization is probably an important mechanism in argon discharges, and in this connection the results of Weissler seem to be confirmed.

ACKNOWLEDGMENTS

The authors are greatly indebted to Professor L. B. Loeb for many searching criticisms and comments, and also for his kind encouragement. They also wish to thank Mr. F. R. Perry for his support of the work and Sir Arthur Fleming, Director, and Mr. B. G. Churcher, Manager, of Metropolitan-Vickers Research Department, for permission to publish the paper.

REFERENCES

- BENNETT, E., 1914, *Trans. Amer. Inst. Elect. Engrs.*, **33**, 571.
 BOULIND, H. F., 1934, *Phil. Mag.*, **18**, 909 ; 1935, *Ibid.*, **20**, 68.
 BOWLS, W. E., 1938, *Phys. Rev.*, **53**, 293.
 BRUCE, J. H., 1930, *Phil. Mag.*, **10**, 476.
 COBINE, J. D., 1941, *Gaseous Conductors* (New York : McGraw-Hill).
 CRAGGS, J. D., and MEEK, J. M., 1943 a, *Nature, Lond.*, **152**, 386 ; 1943 b, *Phys. Rev.*, **64**, 249.
 DRUYVESTEYN, M. J., and PENNING, F. M., 1940, *Rev. Mod. Phys.*, **12**, 87.
 HALE, D. H., 1938, *Phys. Rev.*, **54**, 241 ; 1939, *Ibid.*, **56**, 815.
 HUXLEY, L. G. H., 1928, *Phil. Mag.*, **5**, 721 ; 1930, *Ibid.*, **10**, 185.
 KIP, A. F., 1938, *Phys. Rev.*, **54**, 139 ; 1939, *Ibid.*, **55**, 549.
 KRUTHOF, A. A., and PENNING, F. M., 1936, *Physica*, **3**, 515 ; 1937, *Ibid.*, **4**, 430.
 LOEB, L. B., 1939, *Fundamental Processes of Electrical Discharge in Gases* (New York : John Wiley).
 MASCH, K., 1932, *Arch. Elektrotech.*, **32**, 589.
 MORTON, P. L., 1946, *Phys. Rev.*, **70**, 358.
 PENNING, F. M., 1931, *Phil. Mag.*, **11**, 961.
 PENNING, F. M., MOUBIS, J., and ADDINK, C. C. J., 1933, *Physica*, **13**, 209.
 POSIN, D. Q., 1936, *Phys. Rev.*, **50**, 650.
 SCHULZE, W., 1932, *Z. Phys.*, **78**, 92.
 SLOANE, R. W., 1937, *Phil. Mag.*, **23**, 534.
 TRICHEL, G. W., 1938, *Phys. Rev.*, **54**, 1078 ; 1939, *Ibid.*, **55**, 382.
 TYKOCINER, J. T., TARPLEY, R. E., and PAINE, E. B., 1935, *Bulletin 278, Illinois Univ. Engng. Exptl. Station*.
 VALLE, G., 1926, *Phys. Z.*, **27**, 473.
 WEISSLER, G. L., 1943, *Phys. Rev.*, **63**, 96.

The Pressures Exerted by an Underwater Explosion Bubble on a Circular Target Plate in a Disc-like Baffle

By A. R. BRYANT

Admiralty

MS. received 18 December 1947

ABSTRACT. The general equations of motion are derived for a target plate in a finite rigid baffle under the influence of pressures due to a "point source" on the axis of the target. The fluid is assumed incompressible so that the theory should be a fair approximation in the case where the pressures are due to the oscillation of the gaseous products of an underwater explosion.

The effect of the water loading on the target plate due to its own motion appears in the equation of motion as an increase in the inertia of the target. A formula for this added mass is derived for a circular target plate dishing "proportionally". The added mass is calculated for a target radius half that of the baffle in both the "piston" and "paraboloid" types of proportionally constrained motion, and it is found that it differs from the corresponding value when the baffle is infinite by not more than 15% in both cases.

The pressure distribution over a rigid disc-like baffle is computed for various distances of the explosive source. Except when the source distance is small compared to the baffle radius, the pressures are considerably lower than in the case of an infinite baffle.

§ 1. INTRODUCTION

EXPERIMENTAL studies of the damaging effect of underwater explosions frequently make use of a circular air-backed target plate mounted in a rigid annular baffle. Under the impact of the primary shock-wave sent out by the explosion the target is dished, drawing energy from the incident wave, and the waves diffracted from the baffle, in a complicated manner. The length of the pressure pulse and the distance of the explosive charge from the target are commonly of the same order as the radius of the target plate or baffle, and the extremely difficult diffraction problem has so far only been tackled by very approximate methods.

It is known that under certain circumstances a second pressure pulse is generated by the oscillating bubble formed by the gaseous products of the explosion, which can also contribute in marked degree to the dishing of the target plate. This secondary pulse is not shock-fronted, the rise and fall of pressure being much more gradual than in the primary shock-wave. The pulse length is often several times the linear dimensions of the target, so that an incompressible treatment is likely to be a fair approximation. In the following note the problem is simplified to the consideration of a hydrodynamic point source situated on the axis of a rigid disc. The solution obtained is thought to be of wider interest as a limiting case in acoustical problems involving rigid discs or circular vibrating membranes.

§ 2. STATEMENT OF PROBLEM

Figure 1 shows diagrammatically the rigid disc or baffle AB, centre O, and the explosion centre E. The dotted portion CD represents the light air-backed target plate which acquires a velocity normal to the plane AB under the influence

of the explosion pressures. In practical targets there must always be some rigid structure behind the central portion CD to keep the fluid pressures from the back of the target plate. For the present purpose this is simplified by the assumption that the normal fluid velocity must be zero over the whole of the rear face (uppermost in figure 1) of the disc, and over the portion of the front face outside the target area COD. Complete symmetry about the axis Ox of the disc is assumed. As usual in such problems

the case of "proportional motion" of the target plate is assumed in which each point in the target plate moves with a normal velocity bearing a ratio to the velocity of its centre which is independent of time. The special case of a piston-like target plate in which all points move at the same velocity will be considered in detail.

When steel target plates dish under explosive loading the energy absorbed in the plastic stretching of the steel is usually large compared to the elastic energy involved, and may be taken as proportional to the square of the central deflection. This central deflection will be denoted by z .

Figure 1 also shows the two coordinate systems used. Cylindrical coordinates x, ρ are taken with the origin at the centre of the disc and the positive x direction is the line OE; ρ denotes the distance of any point from this axis.

Oblate spheroidal coordinates r, s are also used, chosen so that the system of ellipsoids $r = \text{constant}$ varies between the limiting case $r = 0$ which reduces to the disc AOB, and $r = \infty$ which is a sphere at infinity. The surfaces $s = \text{constant}$ are hyperboloids of one sheet and vary from $s = -1$ which is the negative Ox axis, through $s = 0$ which is the plane of the disc, to $s = +1$ which is the positive Ox axis. Both the r and s coordinates are therefore continuous in the space lying outside the disc, i.e. in the fluid, negative values of s corresponding to the region behind the disc. A point source of strength 4π is situated at E, which has spheroidal coordinates $(r_0, 1)$ and cylindrical coordinates $(x_0, 0)$. The radius of the disc is C , while the radius of the target plate need not for the moment be specified.

The relation between the two systems of coordinates is

$$x = Crs; \quad \rho = C(r^2 + 1)^{1/2}(1 - s^2)^{1/2}. \quad \dots\dots(1)$$

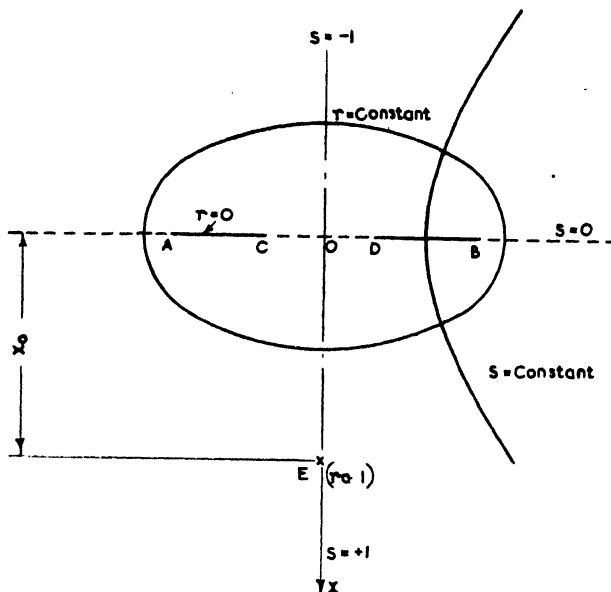


Figure 1. Spheroidal coordinates.

§3. VELOCITY POTENTIALS

Let ϕ_1 be the velocity potential due to a point source of strength 4π at E, in the presence of the rigid disc. Thus ϕ_1 reduces to the reciprocal of the distance at points very close to the source, and produces zero normal velocity over the whole of the disc. Let ϕ_2 be the velocity potential which would arise if the point source were absent, due to the motion of the target plate with unit central velocity in the direction of z increasing; i.e. away from the explosion centre.

The velocity potential due to an explosion bubble of radius a at E will approximate very closely to $a^2\dot{a}\phi_1$ provided that a is small compared to the dimensions and distance of the disc, and this assumption will be made in what follows.

Thus the complete velocity potential for the explosion bubble and the moving target plate is

$$\phi = a^2\dot{a}\phi_1 + \dot{z}\phi_2. \quad \dots\dots(2)$$

§4. LAGRANGIAN ANALYSIS

The system comprising the target plate and the fluid will now be treated by the Lagrangian method of generalized coordinates; the two coordinates necessary to specify the motion of the system are here z and a . Then if T be the kinetic energy of the system, the Lagrangian equations of motion yield

$$d(\partial T/\partial \dot{z})/dt - \partial T/\partial z = Q, \quad \dots\dots(3)$$

and a similar equation for the variable a , which is here irrelevant. Q is the "generalized force" corresponding to the displacement z . It is defined so that if the target plate suffers a small deflection dz the potential energy of the system decreases by $Q dz$. In this case Q is equal to $-kz$, where the constant k may be evaluated from the relation between energy absorbed and deflection z for a particular type of target plate.

By a theorem in hydrodynamics the kinetic energy of the fluid is given by the expression $-\frac{1}{2}\sigma \int \phi \frac{\partial \phi}{\partial n} ds$, where σ is the fluid density taken over all the boundaries of the fluid, i.e. over the disc and bubble surface; the integral vanishes over the boundary at infinity. $\partial \phi/\partial n$ is the gradient of ϕ along a normal drawn inwards towards the fluid, and is therefore the velocity along the outward drawn normal.

To this must be added the kinetic energy of the target plate $\frac{1}{2}m\dot{z}^2$. For a piston, m is its actual mass, while for any other case of proportional motion m may be readily obtained by integration. The total kinetic energy T is therefore given by the expression

$$T = \frac{1}{2}m\dot{z}^2 - \frac{1}{2}\sigma \left\{ a^2\dot{a}^2 \int \phi_1 \frac{\partial \phi_1}{\partial n} ds + a^2\dot{a}\dot{z} \int \phi_1 \frac{\partial \phi_2}{\partial n} ds + a^2\dot{a}\dot{z} \int \phi_2 \frac{\partial \phi_1}{\partial n} ds + \dot{z}^2 \int \phi_2 \frac{\partial \phi_2}{\partial n} ds \right\}, \quad \dots\dots(4)$$

in which the integrals are to be taken over the disc and the bubble. Equation (4) may be simplified by the following considerations:

(i) $\partial \phi_1/\partial n$ vanishes over the disc. Over the bubble the first integral reduces to $-4\pi a$.

(ii) Since ϕ_1 and ϕ_2 are potentials satisfying Laplace's equation, by Green's theorem the second and third integrals are equal.

(iii) The second and fourth integrals vanish over the bubble surface since $\partial\phi_2/\partial n$ has equal and opposite values at corresponding points on opposite sides of the bubble, while ϕ_1 and ϕ_2 are substantially constant over the bubble surface.*

Equation (4) now becomes

$$T = \frac{1}{2}m\dot{z}^2 + 2\pi\sigma a^2\dot{a}^2 - \frac{1}{2}\sigma\dot{z}^2 \int \phi_2 \frac{\partial\phi_2}{\partial n} ds - \sigma a^2\dot{a}\dot{z} \int \phi_1 \frac{\partial\phi_2}{\partial n} ds, \quad \dots\dots(5)$$

in which the integrals are taken over the target only.

The Lagrangian equation (3) now gives the equation of motion of the target

$$\ddot{z}[m + M] + kz = \sigma Pd(a^2\dot{a})/dt, \quad \dots\dots(6)$$

where $M = -\sigma \int \phi_2 \frac{\partial\phi_2}{\partial n} ds$ and $P = \int \phi_1 \frac{\partial\phi_2}{\partial n} ds$ over the target. $\dots\dots(7)$

It may be remarked at this stage that equation (6) would be obtained for any shaped target and rigid baffle, the only difference being in the form of the two potential functions ϕ_1 and ϕ_2 . The quantity M is usually called the "added mass" or "virtual mass", and is the increase of the effective inertia m of the target due to the mass of fluid which is set in motion by it. This added mass is independent of the system of sources which ultimately causes the motion of the plates, but it does depend on the shape and size of the rigid baffle.

In the special case of a piston-like target the two quantities M and P become

$$M' = -\sigma \int \phi_2 ds; \quad P' = \int \phi_1 ds \text{ over the piston.} \quad \dots\dots(7a)$$

In this case it can be seen that the right-hand side of (6) is simply the total load which would be exerted by the source over the region of the disc occupied by the target if the target were not moving.

Equation (6) is similar to the simple equation for the behaviour of a piston on a spring subjected to a pressure varying in a known way with the time. In order to make it of practical use the potentials ϕ_1 and ϕ_2 must now be considered in detail and the quantities M and P evaluated.

§ 5. POTENTIAL DUE TO POINT SOURCE AND DISC

Consider the potential

$$\phi = \sum_{m=0}^{\infty} \frac{i}{C} (2m+1) P_m(ir) Q_m(ir_0) P_m(s), \quad \dots\dots(8)$$

for $r < r_0$ where P_m and Q_m are Legendre functions of the first and second kind respectively, which may be shown to satisfy Laplace's equation when expressed in terms of the spheroidal coordinates. Along the axis of the disc, which is the line $s = +1$, it reduces to

$$\phi = \sum_{m=0}^{\infty} \frac{i}{C} (2m+1) P_m(ir) Q_m(ir_0) = \frac{1}{C(r_0 - r)} = \frac{1}{x_0 - x},$$

this being Heine's expansion. Consequently (8) is the potential of a point source of strength 4π situated at $(r_0, 1)$ in an infinite fluid expressed in the spheroidal coordinates.

* In the limiting case when a tends to zero, ϕ_1 and ϕ_2 may be removed from the integrals; the integrals then become the flux of fluid due to the target's motion through the volume occupied by the bubble, and this clearly tends to zero as a tends to zero.

It follows that the velocity potential due to the point source in the presence of the rigid disc is

$$\phi_1 = \sum_{m=0}^{\infty} \left\{ \frac{i}{C} (2m+1) Q_m(ir_0) P_m(ir) P_m(s) + \frac{2}{\pi C} (4m+3) Q_{2m+1}(ir_0) Q_{2m+1}(ir) P_{2m+1}(s) \right\} \dots\dots (9)$$

for $r < r_0$,

since ϕ_1 in (9) has a zero derivative with respect to r at $r=0$, i.e. at points on the surface of the disc, and tends to the reciprocal of the distance in the neighbourhood of the source.

The case where $r > r_0$ is not required here. The value of ϕ_1 is required in (7) and (7a) for points on the disc, i.e. for $r=0$. It is shown in the Appendix that this expression may be put into finite terms, and the value of ϕ_1 at points on the disc is given by the expression

$$\phi_1 = \frac{1}{C(r_0^2 - s^2 + 1)^{\frac{1}{2}}} \left[1 + \frac{2}{\pi} \cot^{-1} \frac{(r_0^2 - s^2 + 1)^{\frac{1}{2}}}{s} \right], \dots\dots (9a)$$

where the angle $\cot^{-1}[(r_0^2 - s^2 + 1)^{\frac{1}{2}}/s]$ is numerically less than $\pi/2$ and is positive for s positive, i.e. for points in the front of the disc, and negative for points behind the disc where s is negative.

§ 6. POTENTIAL DUE TO MOTION OF THE TARGET PLATE

It will be assumed that the motion of the target plate is axially symmetrical. Let the normal velocity v of a point on the target plate be related to its spheroidal coordinate s by the expression

$$v = \dot{z} \partial \phi_2 / \partial n = f(s) \cdot \dot{z}; \dots\dots (10)$$

$f(s)$ will, of course, be zero for values of s corresponding to points on the disc but outside the target area. For any physically reasonable distribution of target plate velocity the following expansion will be possible:

$$f(s) = s^{-1} \sum_{n=0}^{\infty} a_n P_n(s), \dots\dots (11)$$

where the coefficients a_n are given by

$$a_n = \frac{1}{2} (2n+1) \int_{-1}^{+1} s f(s) P_n(s) ds. \dots\dots (12)$$

The velocity potential ϕ_2 due to the motion of the target plate will then be

$$\phi_2 = -iC \sum_{n=0}^{\infty} \frac{a_n Q_n(ir) P_n(s)}{Q'_n(i0)}, \dots\dots (13)$$

since the normal velocity at the surface of the disc $r=0$ is $(1/Cs)(\partial \phi_2 / \partial r)_{r=0}$. If the target plate moves as a piston, so that the normal velocity is a constant over the central region of the disc or baffle up to some radius, say R , and is zero outside this radius, we have

$$f(s) = 1, \quad S < s < 1; \quad f(s) = 0, \quad -1 < s < S$$

where $S = (1 - R^2/C^2)^{\frac{1}{2}}$,

so that
$$a_n = \frac{1}{2} (2n+1) \int_S^1 s P_n(s) ds. \dots\dots (14)$$

§ 7. DISTRIBUTION OF POTENTIAL ϕ_1
AND PRESSURE OVER DISC

The expression (9a) for ϕ_1 also gives the spatial distribution of pressure over the rigid disc. For a source of strength A at a distance x_0 from the disc the pressure distribution over the disc is given by

$$p = \frac{\sigma \dot{A}}{4\pi (x_0^2 + \rho^2)^{\frac{1}{2}}} \left[1 + \frac{2}{\pi} \cot^{-1} \left(\frac{x_0^2 + \rho^2}{C^2 - \rho^2} \right)^{\frac{1}{2}} \right] = \frac{\sigma \dot{A}}{4\pi} \phi_1. \dots\dots (15)$$

If the disc caused no alteration in the flow the pressure at its centre would be $\sigma \dot{A}/4\pi x_0$, and this may be likened to the "incident pressure" in the acoustic case. In figure 2 the pressure p is divided by this "incident pressure" and

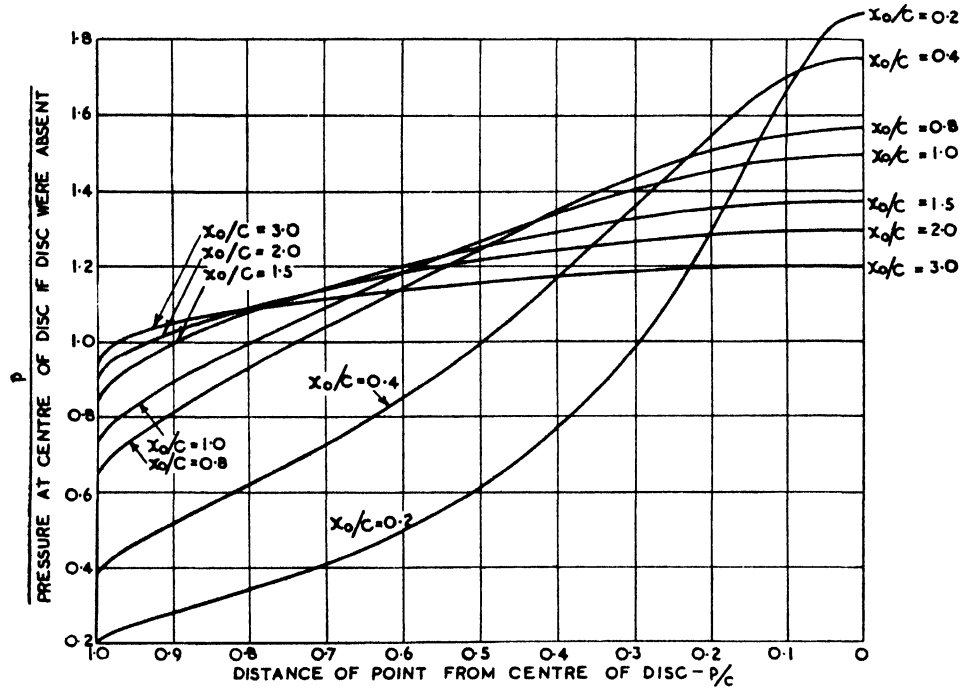


Figure 2. Total pressure p at a point on a rigid disc radius C . Distance of point source along axis of disc $= x_0$.

plotted as a function of distance ρ from the centre of the disc. This quantity can be seen to be equal to $x_0\phi_1$, which is tabulated in table 1. For points in the

Table 1. The quantity $x_0\phi_1$ for a point on the surface of a rigid disc

$x_0/C \rightarrow$	0.2	0.4	0.8	1.0	1.5	2.0	3.0
ρ/C							
0.0	1.874	1.758	1.571	1.500	1.374	1.295	1.205
0.1409	1.508	1.643	1.539	1.479	1.364	1.289	1.201
0.2431	1.144	1.464	1.480	1.439	1.344	1.276	1.195
0.3122	0.951	1.330	1.427	1.402	1.325	1.264	1.188
0.4359	0.704	1.102	1.314	1.319	1.280	1.235	1.171
0.600	0.498	0.850	1.144	1.186	1.200	1.181	1.140
0.7141	0.400	0.709	1.021	1.082	1.132	1.133	1.111
0.800	0.340	0.615	0.927	0.999	1.073	1.089	1.084
0.9165	0.268	0.497	0.791	0.872	0.975	1.013	1.034
0.9798	0.225	0.423	0.696	0.779	0.897	0.949	0.989
1.0	0.196	0.371	0.625	0.707	0.832	0.894	0.949

rear of the disc the second term in the square brackets of (15) should be given a negative sign.

§ 8. EVALUATION OF THE PRESSURE FUNCTION P

When the distribution of velocity over the target plate $\partial\phi_2/\partial n$ is known, the pressure function P appearing in the equation of motion is obtained from equations (9a) and (10):

$$P = \int \phi_1 \frac{\partial\phi_2}{\partial n} ds = 2\pi C \int \frac{sf(s)}{(r_0^2 - s^2 + 1)^{\frac{1}{2}}} \left[1 + \frac{2}{\pi} \cot^{-1} \frac{(r_0^2 - s^2 + 1)^{\frac{1}{2}}}{s} \right] ds, \quad \dots (16)$$

the integral to be taken over the target plate. In the case of a target plate of radius R , which has a constant velocity of unity over its surface, the value of P is

$$P' = 2\pi \left[(x_0^2 + R^2)^{\frac{1}{2}} - x_0 \right] - 4(x_0^2 + C^2)^{\frac{1}{2}} \left[\sin \theta - \theta \cos \theta \right]_{\theta_1}^{\theta_2}, \quad \dots (17)$$

where $\theta_1 = \cot^{-1} x_0/C$, $\theta_2 = \cot^{-1} [(x_0^2 + R^2)/(C^2 - R^2)]^{\frac{1}{2}}$, C being the disc or baffle radius, and x_0 the distance of the point source. The ratio of the average driving pressure over the piston (i.e. the right-hand side of (6)) to the so-called "incident pressure" is seen to be $x_0 P'/\pi R^2$; this quantity is plotted in figure 3 and given in table 2 for a range of values of x_0/C and R/C .

Table 2. The quantity $x_0 P'/\pi R^2$ for a piston-like target

$x_0/C \rightarrow$	0.2	0.4	0.8	1.0	1.5	2.0	4.0
R/C							
↓							
0.0	1.874	1.758	1.570	1.500	1.374	1.295	1.156
0.1409	1.676	1.704	1.556	1.488	1.369	1.295	1.156
0.2431	1.427	1.600	1.532	1.469	1.359	1.285	1.153
0.3122	1.275	1.515	1.497	1.450	1.350	1.280	1.151
0.4359	1.049	1.367	1.434	1.406	1.325	1.265	1.145
0.6000	0.831	1.177	1.336	1.333	1.286	1.238	1.134
0.7141	0.718	1.059	1.261	1.274	1.250	1.214	1.125
0.9165	0.565	0.877	1.121	1.157	1.174	1.160	1.101
1.0	0.513	0.808	1.057	1.102	1.134	1.130	1.087

Figures 2 and 3 show that the principal factor in determining the increase in pressure over the target area is the ratio of the baffle radius to the source distance. Increasing the baffle round a given target plate makes an appreciable increase in the pressures over the target only when the baffle radius becomes comparable with the source distance. The pressures over the target only approach the value obtained with an infinite baffle when the baffle radius is large compared with the source distance.

§ 9. EVALUATION OF THE ADDED MASS M

Using equations (7), (10), (11) and (13), the added mass becomes

$$\begin{aligned} M &= \sigma \int \phi_2 \frac{\partial\phi_2}{\partial n} ds = \sigma \int_{-1}^1 iC^3 \sum_{n=0}^{\infty} \frac{a_n Q_n(i0) P_n(s)}{Q'_n(i0)} f(s) 2\pi s ds \\ &= 4\pi\sigma iC^3 \sum_{n=0}^{\infty} \frac{a_n^2 Q_n(i0)}{(2n+1)Q'_n(i0)}. \quad \dots (18) \end{aligned}$$

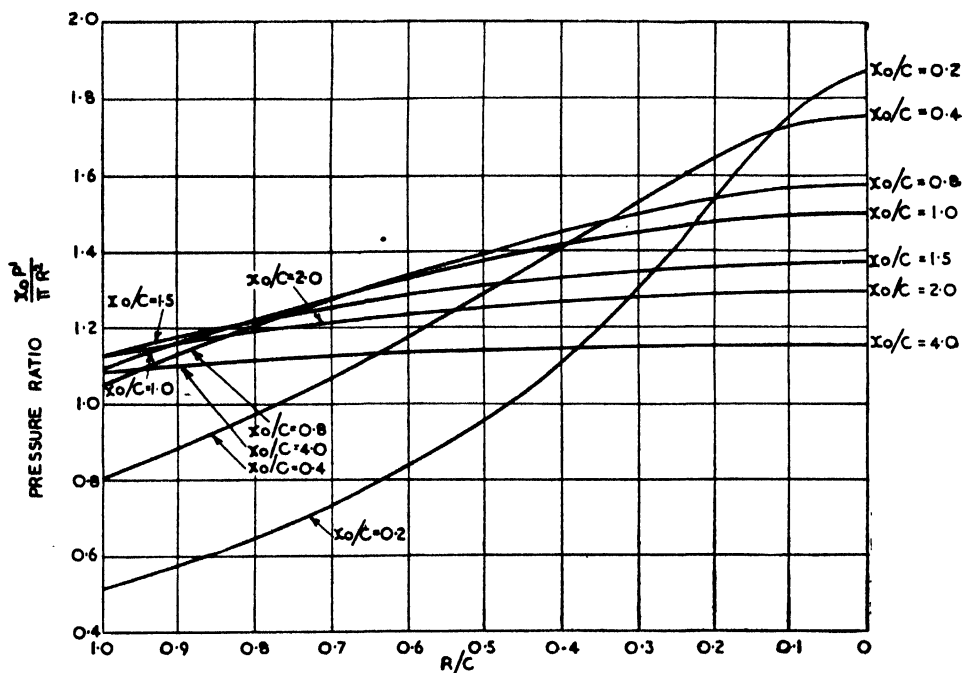


Figure 3. Average pressure over piston radius R in baffle radius C divided by pressure produced by charge at piston centre in absence of piston.

§ 10. ADDED MASS FOR PISTON-LIKE TARGET

As a numerical example consider a piston-like target whose radius R is one half the radius of the baffle C . The coefficients a_n have been computed as far as a_9 . Though the series is rather slowly convergent, the sum to nine terms is estimated to be within two or three per cent of the sum to infinity and yields for the added mass M' the value $2.36\sigma R^3$. This should be compared with a piston in an infinite baffle when the added mass is $2.667\sigma R^3$, i.e. about 12% greater.

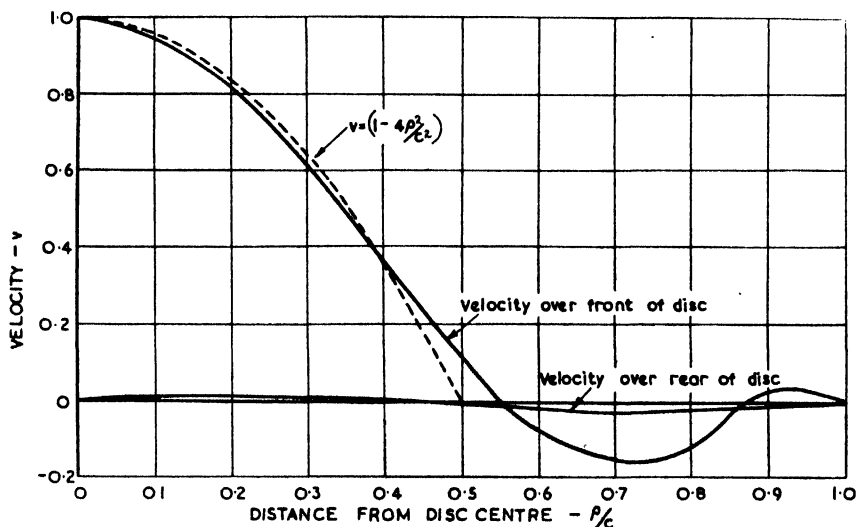


Figure 4. Approximation to parabolic distribution of velocity over target used in "added mass" calculation.

§ 11. ADDED MASS FOR PARABOLOIDAL VELOCITY DISTRIBUTION

The full line in figure 4 represents an assumed velocity distribution using only the seven coefficients a_0 to a_6 which were determined by a numerical method. The velocity distribution is a fair approximation to a parabola for a target plate whose radius is half that of the baffle—the exact parabola is shown dotted in figure 4 for comparison.

Using this assumed velocity distribution, the added mass is calculated to be $0.707\sigma R^3$. This is to be compared with the figure $0.813\sigma R^3$ for a parabolic distribution in an infinite baffle.

It appears, therefore, that for target plates whose radius is not greater than half the baffle radius, the added mass for both the "piston" and "paraboloid" types of dishing is within 10 to 15% of the added masses in the case where the baffle is infinite.

ACKNOWLEDGMENT

The work described above was carried out at the Road Research Laboratory of the Department of Scientific and Industrial Research, under the direction of the Director of Road Research.

APPENDIX

Reduction of the potential expression to finite terms

The first term in equation (9) for ϕ_1 , the potential due to the point source and rigid disc, is simply the reciprocal of the distance between the point (r, s) and the source at $(r_0, 1)$, and may be written down in finite terms at once. This relation is used as the starting point for the reduction of the second term in (9). Thus using (1):

$$i \sum_{m=0}^{\infty} (2m+1) Q_m(ir_0) P_m(ir) P_m(s) = [(r_0 - rs)^2 + (1+r^2)(1-s^2)]^{-\frac{1}{2}} \dots\dots (i)$$

$$= (r_0 s + i\alpha - r)^{-\frac{1}{2}} (r_0 s - i\alpha - r)^{-\frac{1}{2}}. \dots\dots (ii)$$

In (i) and (ii) $-1 < s < 1$, s real, and α is the positive root of $(r_0^2 + 1)^{\frac{1}{2}}(1-s^2)^{\frac{1}{2}}$. The left-hand side of (i) is uniformly convergent with respect to r for all r_0 except r_0 imaginary and between ± 1 and for all r lying within an ellipse in the Argand diagram having its foci at ± 1 and passing through the point r_0 . In particular it is convergent for all real r_0 if r is imaginary and between $\pm i$.

Now if z and r be any complex numbers (z not real and between ± 1) Neumann's expression for $Q_m(z)$ is

$$Q_m(z) = \frac{1}{2} \int_{-1}^1 \frac{P_m(r) dr}{z - r}. \dots\dots (iii)$$

Putting $ir' = r$ and $iz' = z$, and then omitting primes, we get

$$Q_m(z) = \frac{1}{2} \int_i^{-i} \frac{P_m(ir) dr}{z - r}, \dots\dots (iv)$$

provided now that z is not a pure imaginary between $\pm i$.

Now multiply both sides of (ii) by $\frac{1}{2}(z-r)^{-1}$ and integrate with respect to r from i to $-i$. Reversal of the order of summation and integration in order to integrate term by term is justified because of the uniformity of convergence, and this gives, with (iv),

$$i \sum_{m=0}^{\infty} (2m+1) Q_m(ir_0) Q_m(iz) P_m(s) = \frac{1}{2} \int_i^{-i} \frac{dr}{(z-r)(\lambda-r)^{\frac{1}{2}}(\lambda'-r)^{\frac{1}{2}}}, \quad \dots\dots (v)$$

where $\lambda = r_0 s + i\alpha$, $\lambda' = r_0 s - i\alpha$.

By changing the variable in the integral in (v) to $x = z - r$, the integration may be performed and (v) becomes

$$i\Sigma = \frac{1}{2(\lambda-z)^{\frac{1}{2}}(\lambda'-z)^{\frac{1}{2}}} \left\{ \sinh^{-1} \left[\frac{r_0 s - z}{\alpha} + \frac{(\lambda-z)(\lambda'-z)z}{\alpha(z^2+1)} - \frac{i(\lambda-z)(\lambda'-z)}{\alpha(z^2+1)} \right] \right. \\ \left. - \sinh^{-1} \left[\frac{r_0 s - z}{\alpha} + \frac{(\lambda-z)(\lambda'-z)z}{\alpha(z^2+1)} + \frac{i(\lambda'-z)(\lambda'-z)}{\alpha(z^2+1)} \right] \right\}. \quad \dots\dots (vi)$$

The expression on the right of (vi) is thus of the form

$$\frac{1}{2X} \{ \sinh^{-1}(A - iB) - \sinh^{-1}(A + iB) \},$$

where X , A , and B are all positive and real for r_0 , s , z , positive and real and $z < r_0 s$. Now rewrite equation (vi) with $-s$ in place of s . Since the P_m functions are even or odd as m is even or odd the effect of this is to reverse the sign of all the odd order terms in the summation and to leave the even order terms unaffected. If now this modified expression be subtracted from (vi) the even order terms will disappear and each odd order term will appear twice. Hence

$$2i \sum_{m=0}^{\infty} (4m+3) Q_{2m+1}(ir_0) Q_{2m+1}(iz) P_{2m+1}(s) \\ = \frac{1}{2X} \{ \sinh^{-1}(A - iB) - \sinh^{-1}(A + iB) \} \\ - \frac{1}{2X'} \{ \sinh^{-1}(A' - iB') - \sinh^{-1}(A' + iB') \}, \quad \dots\dots (vii)$$

where X' , A' and B' are the expressions obtained by reversing the sign of s in X , A and B respectively. We now require the limit of (vii) as z tends to zero. This is

$$i \sum_{m=0}^{\infty} (4m+3) Q_{2m+1}(ir_0) Q_{2m+1}(i0) P_{2m+1}(s) \\ = \frac{1}{4F} \{ \sinh^{-1}(D - iE) - \sinh^{-1}(D + iE) \} \\ - \{ \sinh^{-1}(-D - iE) - \sinh^{-1}(-D + iE) \}, \quad \dots\dots (viii)$$

where $D = r_0 s / \alpha$; $E = (r_0^2 + 1 - s^2) / \alpha$; $F = (r_0^2 + 1 - s^2)^{\frac{1}{2}}$. D , E and F are thus real and positive for s positive.

The expression on the left of (viii) is $C\pi i/2$ times the second term in the expression for the pressure on the disc—it represents the increase in the pressure

due to the source caused by the alteration of the flow round the rigid disc. Since this solution must correspond to physical reality the following two statements may be made:

(a) The right-hand side of (viii) must be a pure imaginary, positive when s is positive.

(b) The term within the second square brackets in (viii) must not be *numerically equal* to the first term. For if it were equal in magnitude and sign the expression on the right-hand side of (viii) would vanish, which is physically impossible. If, on the other hand, it were equal in magnitude, but of opposite sign, then the expression as a whole would be an even function of s , implying that the added pressure due to the presence of the disc is of the same sign on both sides of the disc, which is again physically incorrect, since the effect of any rigid obstacle is, in general, to increase the pressure on the side facing the source, and to decrease it on the other side.

It will be observed that the many-valued functions $\sinh^{-1}$ and Q_m appear in equation (viii) and in the expression (9) for the potential. It is clear, therefore, that the particular branch of each function must be chosen so that the solution corresponds with the physical reality. In particular conditions (a) and (b) above enable one to choose the branch of the $\sinh^{-1}$ function according to the following rule. Let μ be a positive number between zero and infinity, and let η be a real angle between 0 and $\pi/2$. We choose that branch of the $\sinh^{-1}$ function in which the real part is always positive, and the imaginary part always lies between $-\pi$ and $+\pi$. If then $\mu + i\eta = \sinh^{-1}(D + iE)$ where D and E are real and positive, the following scheme covers all possible cases and gives the correct value for the branch we have chosen.

$$\begin{aligned}\sinh^{-1}(D + iE) &= \mu + i\eta; & \sinh^{-1}(-D + iE) &= \mu - i\eta + i\pi, \\ \sinh^{-1}(D - iE) &= \mu - i\eta; & \sinh^{-1}(-D - iE) &= \mu + i\eta - i\pi.\end{aligned}$$

It follows that

$$\sum_{m=0}^{\infty} (4m+3)Q_{2m+1}(ir_0)Q_{2m+1}(i0)P_{2m+1}(s) = \frac{1}{(r_0^2 - s^2 + 1)^{\frac{1}{2}}} \left[\frac{\pi}{2} - \theta \right], \quad \dots\dots (ix)$$

where θ is the angle between 0 and $\pi/2$ which is the imaginary part of

$$\sinh^{-1}(r_0 s/\alpha + i(r_0^2 + 1 - s^2)/\alpha).$$

On evaluating this imaginary part it is found to be the angle $\tan^{-1}[(r_0^2 + 1 - s^2)^{\frac{1}{2}}/s]$. Thus finally the potential at the point $(0, s)$ on the disc is

$$\phi_1 = \frac{1}{C(r_0^2 + 1 - s^2)^{\frac{1}{2}}} \left[1 + \frac{2}{\pi} \cot^{-1} \frac{(r_0^2 + 1 - s^2)^{\frac{1}{2}}}{s} \right], \quad \dots\dots (x)$$

where the angle $\cot^{-1}[(r_0^2 + 1 - s^2)^{\frac{1}{2}}/s]$ is numerically less than $\pi/2$ and is positive or negative as s is positive or negative, i.e. for points in front of or behind the disc respectively.

The Impact of a Solid on a Liquid Surface

By E. G. RICHARDSON

King's College, Newcastle upon Tyne

MS. received 17 February 1948

ABSTRACT. Part I consists of a study of the motion of a solid sphere when it strikes a liquid surface vertically. The shape of the air cavity formed is delineated and explained in terms of potential flow and the resistance experienced by the sphere in this stage is measured. These factors are found to scale on the basis of a Froude parameter. The pressure developed in the cavity is also measured and found to involve reverberations in the cavity of large amplitude.

In Part II the impact forces on the sphere and other forms of projectile on entry into the liquid are studied and compared with theory. The mechanism and forces involved when a sphere ricochets after entry at oblique angles are also discussed.

GENERAL INTRODUCTION

EXPERIMENTAL work on the impact of solids on water has been in the past carried out with two objects in view: (i) to study the form of the cavity and splash and, in one instance, the sounds associated with the collapse of the cavity; (ii) to measure the force of impact. During the late war interest became centred on the matter from the point of view of the behaviour of projectiles entering water, and from this aspect the author carried out the research here described at the Royal Aircraft Establishment in the period 1942–1946. Though this work was concerned with applications to water-entry ballistics and, to a lesser extent, to the impact of seaplanes on landing, it was planned as a piece of fundamental research using the sphere as basic form of solid.

The work is divided into two parts, in accordance with the classification made above.

PART I. CAVITY FORMATION

§ 1. INTRODUCTION

The pioneer work of Worthington (1897), which surely deserves the epithet of “classical”, was done with steel spheres dropped or shot from an air-gun into water. Mallock (1918), whose paper is entitled “Sounds produced by drops falling on water” (but seems to contain no reference to sound), repeated some of these and gave a tentative explanation of the shapes of the cavities produced. Bell (1924) used other liquids, such as glycerine, but with the ball entering at such low speeds that scarcely any cavity was formed. Ramsauer (1927) shot a pellet into the side of a thin-walled tank and photographed the cavity formed, under somewhat special conditions, since the water surface was not open to the air until the wall was punctured.

All these investigators took single photographs under spark illumination.

The improvement in high-speed ciné-photography permits of a series of photographs being taken at one shot and so of the resistance to motion in the cavity stage being investigated. This, and a study of the evolution of the cavity form, was the main object of the author's experiments.

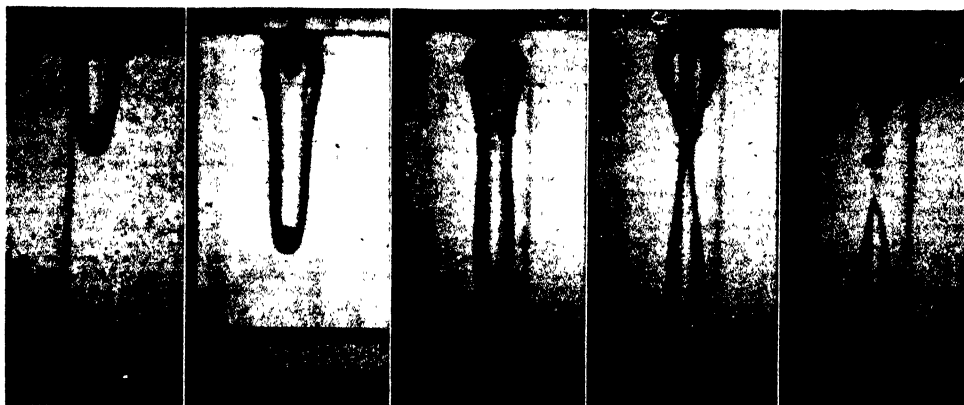


Figure 1. Entry of sphere to water under atmospheric pressure.

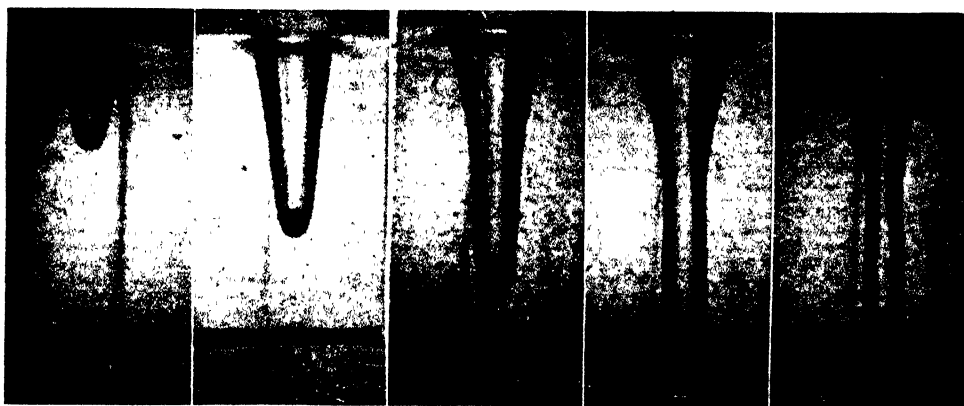


Figure 2. Entry of sphere to water under one-fifth atmospheric pressure.

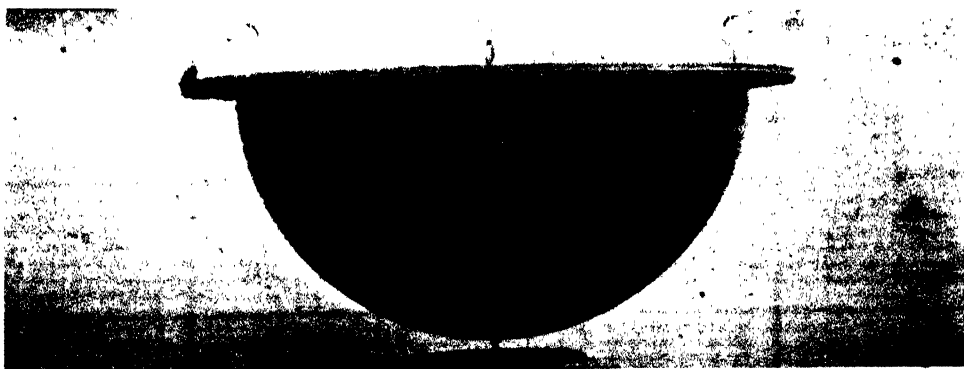


Figure 10. Twenty-inch diameter hemisphere for impact force measurements.

PLATE I.

To face page 352



Figure 3. Details of cavity and splash due to water entry of sphere.

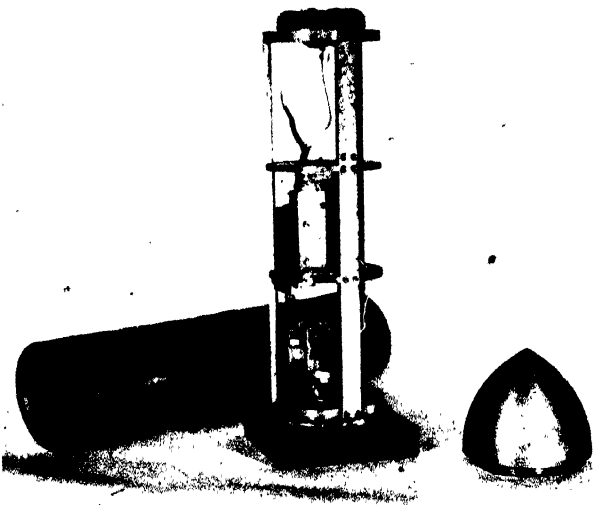
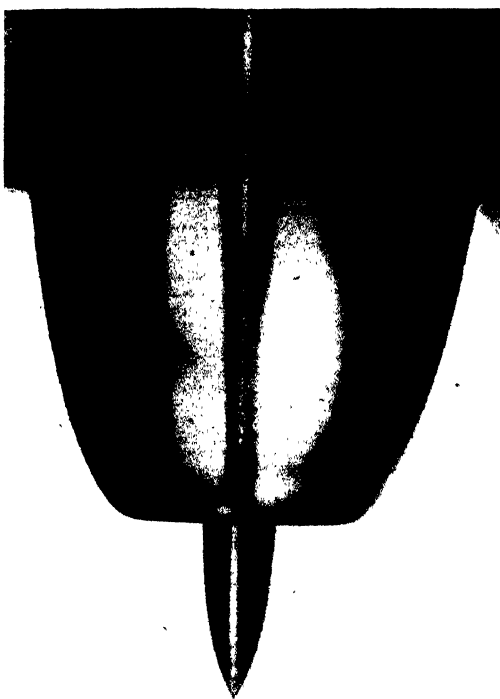


Figure 5. Cavity pressure recorder.



(a)



(b)

Figure 7. Cavity forms behind disc at entry to water (a); in cavitation tunnel (b).

§ 2. EXPERIMENTS

The shots were standard ball bearings, of diameters $\frac{1}{16}$ " to 1", which were projected vertically into the liquid contained in a glass-sided tank. In some instances an air-gun was used, but as it was felt that the expulsion of air following the ball might affect the cavity, the majority of the experiments were done with the projectile falling freely. To get a sufficient fall indoors, a lift-shaft at Farnborough was used and, for greater impact speeds, the interior of the Kew pagoda, which had at the time holes cut in the wooden floors, giving a total drop of 126 ft.

The photography was done with a ciné-camera at 200 frames per sec. Typical prints are shown on plates I (figures 1, 2) and II (figure 3).

§ 3. SHAPE OF THE CAVITY

In order to study the flow of water in detail as the sphere passes vertically downward, a series of fine air bubbles was produced beneath it. They issued under pressure from a fine glass capillary on the end of a tube which was pivoted above the surface and rocked to and fro in the meridian plane of the sphere as it was shot in. The upward speed of the bubbles was very small compared to the downward speed of the projectile.

A pair of arditron flash-lamps were arranged to discharge, one shortly after the other, when the sphere had reached a set depth. These time settings were fixed from a datum instant determined by the breaking of a carbon pencil in the arditron circuits just above the surface of the water as the sphere entered. This caused the double exposure of a photographic plate, exhibiting thus the relative movement of all the bubbles in the field of view and of the sphere itself. Some photographs were taken using aniline drops in place of air bubbles, but these were not suitable for analysis.

From such photographs, and from some very good ones lent to the author by a group of American photographers who used the Egerton stroboscope, the field of flow round the sphere was deduced.

The right-hand side of figure 4 shows by the length of the arrows the velocity field relative to the camera with, as a tentative approximation to the flow, the stream lines due to a source set one-quarter of the radius eccentric to the sphere in the direction of motion. On the left-hand side this flow is compounded with the velocity of the sphere to give, by the length of the arrows, the field relative to it. The potential due to

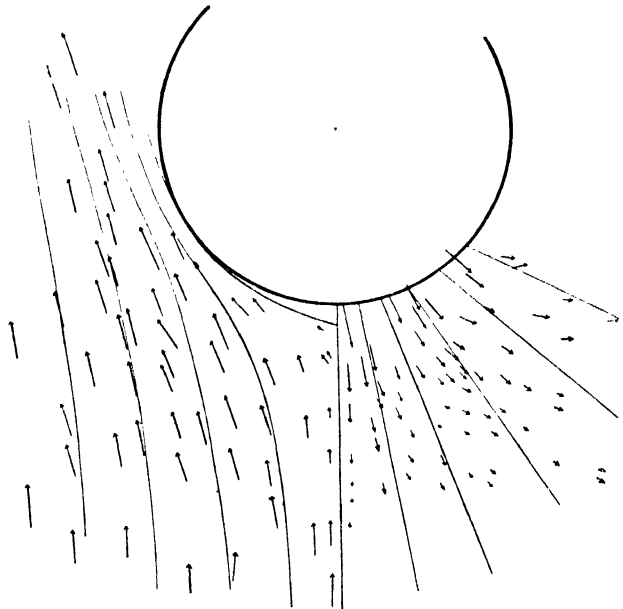


Figure 4. Experimental and assimilated theoretical flow round sphere.

a source may in fact be added to that of a uniform stream to give that at any point r, θ (cf. Bauer 1926) with r measured from the source and θ from the zenith $\phi/V = a^2/r - r \cos \theta$, whence the streamlines $\psi/V = a^2 \cos \theta - \frac{1}{2}r^2 \sin^2 \theta$. The streamline which divides to pass the obstacle is, however, given by $a^2 = a^2 \cos \theta - \frac{1}{2}r^2 \sin^2 \theta$ or $r = a \operatorname{cosec} \theta/2$, so that the source cannot be set to make the dividing line conform to a sphere of radius a . In figure 4 it is set to give the observed value θ_1 for tangential secession from the surface, i.e. 115° .

After the liquid has been cleft by the passage of the sphere it reunites under the action of gravity, so nicking off the cavity behind the sphere. At the instant of closure one can represent the steady state of the cavity by a potential flow due to a source, as in figure 4, together with a linear sink of the same total strength stretching from the source vertically up to a point distant a from the closure. The corresponding stream function for a point distant r from the source and c from the upper end of the sink is $\psi/V = a^2 \cos \theta - \frac{1}{2}r^2 \sin^2 \theta - a(r - c)$.

In fact, the distance l from the centre of the sphere to the point of closure is a function of V , being given, for water, by $V^2/gl = 10$ (see § 5).

It must be borne in mind that the assimilation to a potential flow is valid only for the steady state, and that in making this comparison we are ignoring the deceleration which the sphere is actually suffering. Nevertheless, as figure 4 shows, we can derive a good approximation to the field of flow by this method.

The liquids, other than water, which were used were: zinc chloride solution (specific gravity 1.75), petrol (specific gravity 0.71) and glycerine (specific gravity 1.26). It appears that θ_1 is independent of density. On the other hand, the time of closure is delayed beyond that given by the empirical formula just noted when the solid cleaves a very viscous liquid like glycerine.

Often another closure occurs at the surface (cf. figure 1) before the deep one to which the preceding discussion applies. This is due to the piling up of surface water at first impact, which causes faster closing of the lid of the cavity but, being evanescent, has little effect on the closure at depth. Surface tension forces are comparatively small at the speeds involved in this work, but may have important effects at low speeds, as Worthington and Bell have shown.

Failure to close first at the surface seems to be associated with the ability of the water to shoot up and out in a pronounced splash, and this is favoured by reduction of the atmospheric pressure (figure 2) or, at low speeds, of the surface tension (as in zinc chloride solution).

That the cavity still takes the same shape when the ball hits the water at supersonic speeds may be seen from photographs of Ramsauer and of McMillen (1945), who used rifles to fire shots into water.

§ 4. PRESSURE IN THE CAVITY

An attempt has been made to measure the air pressure developed in the cavity during the under-water trajectory, since such information may help to elucidate the mechanism of cavity closure and disintegration. It is not to be expected that the pressure will deviate much from the atmospheric or the local hydrostatic value, so a sensitive recorder is necessary. For the trials a hollow cylindrical projectile with a heavy ogival nose was constructed. Let into the (plane) base, and flush with it, was a thin steel diaphragm (0.004 in. thick and 1.5 in. diameter) attached by a central boss to a D.V.L. scratch recorder carried inside the shell.

The recorder (Jones and Davies 1934) has a stylus moved by deflection of the diaphragm, which scratches a record of the pressure on a stainless steel ring, rotated by an electric motor. Another stylus inscribes a datum line and time marks ($1/50$ sec.), being actuated by a solenoid to which a current is supplied periodically by clockwork mechanism or by a standard tuning-fork. This time-marker was worked either by a battery carried in the shell or from the fork through "flying" leads. The calibration was effected by covering the base with a capsule to which air pressure, manometrically measured, could be introduced.

Figure 5 (plate II) shows the recorder with head and case removed.

It was necessary to ensure that the records in the cavity were not vitiated by the natural response of the mechanism, excited in seismograph fashion. Two tests were made: first, a drop into water was made with the capsule covering the base and hermetically sealed, when no change from atmospheric pressure could be perceived; secondly, the apparatus was dropped on the ground to record its reaction. This was a single peak which returned to zero without oscillation in about $1/100$ sec. (figure 6(c)). It was therefore considered that the recorder was

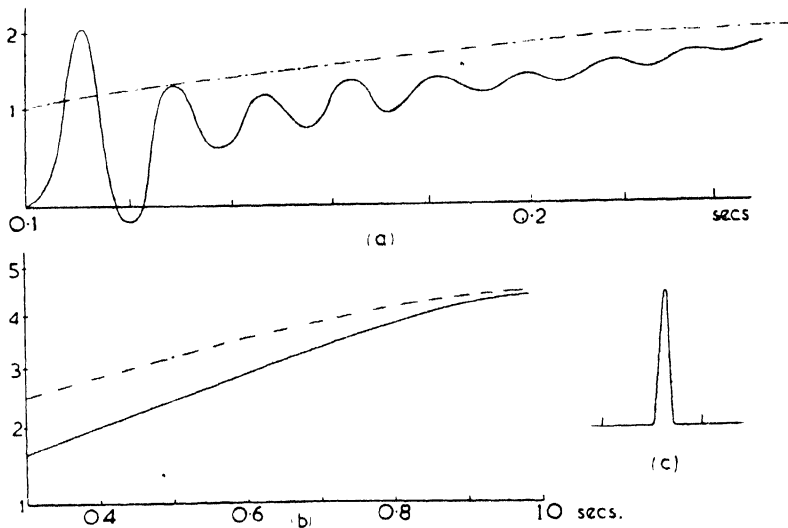


Figure 6. Cavity pressure records shown by continuous line with local hydrostatic pressure shown by chain line (a); continued at (b); record of drop on to ground (c).

sufficiently dead-beat to give a true record of pressure in the lesser deceleration experienced within the cavity when the projectile was dropped into water. The adoption of the ogival nose prevented any disturbance of the trace on hitting the water (Part II, §3). In order to mark "entry" on the records it was arranged that the time marker should come into action just as the nose touched the surface.

The apparatus was dropped nose-first into a large glass-sided tank 20 ft. deep. Ciné-photography at 200 frames/sec. enabled a depth-time record to be kept, and this was converted into one reading local hydrostatic pressure for comparison with the cavity pressure. A typical pair for entry at 20 ft/sec. is shown in figure 6 (a) and its continuation in 6 (b), beginning when the pressure in the cavity (shown by the continuous line) started to deviate from atmospheric. The chain line gives the hydrostatic pressure, calculated from the depth of the projectile.

The following features may be discerned from such records:

(i) The cavity pressure, at this comparatively low entry speed, remains nearly atmospheric until the cavity closes; (ii) as the cavity closes, large oscillations of frequency 20–80/sec. and initial amplitude up to $1/20$ atm. supervene, but are rapidly damped and disappear in about $\frac{1}{10}$ sec., rising in frequency as they do so, perhaps due to loss of air in smaller bubbles; (iii) the mean pressure on the diaphragm rises rapidly from this epoch as the projectile sinks, eventually catching up with the local hydrostatic pressure when the speed has dropped and the cavity mostly bubbled away.

The discovery of these acoustic vibrations seems to have completed Mallock's unfinished task. Their large decrement suggests that they are either "reverberations" of the cavity damped by turbulence, or pulsations of the shell of water surrounding the cavity, as oscillations of figure would be of lower frequency and would not change the internal pressure. On the whole the motions are suggestive of cavity resonance like those of a pop-gun. The decrement has a mean value of 2 per sec., rather more than that in Kirchhoff's theory for this case (which applies to small displacements) but of the same order as those recorded in a pipe of rigid walls by Lehmann (1934) at the corresponding frequency and amplitude.

These powerful but short-lived oscillations, as well as the significant difference (for the time being) between the internal and external pressure at the cavity, must be the cause of its breaking up into individual bubbles and final evanescence.

§ 5. RESISTANCE IN CAVITY

If a sufficiently high frequency of photography is used it is possible to plot the displacement-time curve of the projectile after it has formed a cavity and so calculate its resistance in motion. At the entry speeds involved (4–40 m/sec.) the frictional resistance was the paramount force causing deceleration. If M is the mass of the projectile, d its diameter, V its speed, ρ the density of the liquid,

$$M\partial V/\partial t = Mg - C_D(\pi d^2/4)(\frac{1}{2}\rho V^2), \quad \dots\dots(1)$$

where C_D is a drag coefficient which relates the frictional force to the projected area of the projectile and the pressure at its front stagnation point in the conventional fashion.

This may be written, when σ is the specific gravity of the solid in terms of the liquid and s the path traced by the sphere,

$$\frac{1}{8}\pi d^3\sigma(g - V\partial V/\partial s) = C_D\pi d^2V^2/8.$$

When the speed and deceleration are such that the effect of gravity can be neglected, we have for the drag coefficient

$$C_D = 3.05(d\sigma/s)\log(V_0/V), \quad \dots\dots(2)$$

V_0 and V being the speeds at the two ends of the track s .

In the first experiments, 100 frames/sec. were used in the ciné-camera. This proved inadequate for measuring the drag, and later 200 frames/sec. or 2000 frames/sec. (the latter for shots at 40 m/sec.) were employed. Values of V were estimated by drawing slopes to the displacement-time curves at different points on the trajectory.

An analysis of the single shot at an entry speed of 650 m/sec. recorded by Ramsauer (1929) (using individual spark photographs) was also made and the drag for his sphere included with the present ones in table 1.

Table 1

Media	Steel in water	Steel in water (low press.)	Lead in water	Steel in zinc chloride	Steel in petrol	Steel in glycerine	Steel in water (Ramsauer),
Diameter (cm.)	1-2.5	1-2.5	7	1-2.5	1-2.5	1-2.5	1.1
σ	8	8	12	4.5	11	6.5	8
Mean Reynolds number } C_D	3×10^5	3×10^5	10^6	3×10^4	3×10^5	3×10^2	5×10^3
	0.30	0.30	0.28	0.30	0.34	0.50	0.32
C_D for full immersion }	0.4	0.4	0.4	0.4	0.4	0.7	0.5

As one would expect, the values of C_D are less than those commonly associated with fully immersed spheres (quoted in the final row of the above table from *Modern Developments in Fluid Dynamics*) at these values of Reynolds number, and are unaffected by the approach to the velocity of sound (in Ramsauer's case), whereas at the Reynolds and Mach numbers corresponding to this (mean) speed the drag of a fully immersed sphere would have fallen considerably.

The motion of the projectile after entry appears to be unaffected by air pressure. Some of the experiments were repeated in a large chamber in which the air pressure could be reduced to one-fifth of atmospheric and the firing of the shot and running of the camera controlled from outside; the drag coefficient was unchanged from that at the same Reynolds number and atmospheric pressure.

It is of interest to relate this type of cavity formation due to impact on a free surface to the cavitation which ensues when a submerged body is moved through a liquid at such a speed that a vacuum or, more correctly speaking, a space saturated with the vapour of the liquid, is formed behind the projectile. As the speed increases, this vaporous cavity increases in size until it extends to a great distance in the rear of the body. Measurements in cavitation tunnels, i.e. water channels in which air-free water is circulated at high speed past a model held on a force balance, enable one to compare this type of cavitation, both as to cavity shape and resistance, with that due to impact on a free surface. If the cavitation number is defined as $(p_1 - p_0)/\frac{1}{2}\rho V^2$, where p_1 is the actual pressure at the point on the solid where cavitation ensues and p_0 is the saturation vapour pressure, then photographic comparison indicates that free-surface cavities correspond in shape to those having a zero-cavitation number. On plate II (figure 7 (a), (b)) are shown the cavity form of a disc at entry into water and in a cavitation tunnel, at a cavitation number of 0.66. Measurements of the resistance of a spherically-nosed model in the Haslar cavitation tank at a Reynolds number of 3×10^6 and cavitation number 0.25, the lowest that can be reached therein, indicate that $C_D = 0.2$ in the immersed model. This may be compared with the values of table 1. As closely as one can estimate from the Haslar photographs, the water secedes from the surface of a sphere in the tunnel at the same latitude (65° from the pole) as in free-surface cavities.

A number of attempts have been made to calculate theoretically the drag coefficient of a sphere in a cavity. Most of these interpolate the angle just quoted. A working approximation to the experimental value of 0.3 may be derived by integrating the pressure distribution round a sphere, as observed in wind tunnels, from the pole to the latitude at which it becomes equal to the static pressure (about 45°), then "stretching" this same distribution to make it cover a greater area of the surface to correspond to the wetted portion in the cavity, i. e. to a latitude of 65° . The value of C_D is in fact only 0.27 if the wind-tunnel distribution of pressure to an upper limit of 45° is taken, but becomes 0.37 if this limit is set to 65° .

One can also derive a value for C_D from the flow pattern of figure 4. An application of Bernoulli's theorem leads to the expression

$$(p - p_0)/\rho = \frac{1}{2} V^2 \{ (2a^2/r^2) \cos \theta + a^4/r^4 \},$$

where p is the pressure at any point r, θ in the field and p_0 the static pressure,

whence
$$C_D = \int_{\pi}^{\theta_1} (4 \sin \theta \cos \theta + 2 \sin \theta) d\theta,$$

neglecting the discrepancy between the dividing streamline and the surface of the sphere. This gives 0.40 with $\theta_1 = 115^\circ$.

§ 6. BASIS OF SIMILARITY

For Reynolds numbers between 10^3 and 5×10^6 it appears from (1) that the motion during cavity persistence is given by:

$$g - dV/dt = \frac{3}{4} C_D V^2 / d\sigma.$$

The specific gravity of the solid and the compressibility of gas and liquid (i. e. the imminence of the velocity of sound) have no great effect on the coefficient C_D . Now phenomena in which the forces are functions of V^2/d , such as the motion of ships, scale on a Froude basis, i. e. they are described in terms of the Froude number V^2/gd . Experiments made with spheres of different diameters at different speeds will experience the same forces and have identical cavity forms at corresponding times if the Froude number is the same.

A similarity based on $V^2/(d\partial V/\partial t)$ does, however, neglect one factor, viz. the air pressure. In so far as these pressures in the cavity are functions of the depth below the free surface and the speed with which the air can enter, they are not scaled on a Froude basis. To the order of accuracy of these experiments, however, no change of drag coefficient with air pressure was detectable in a 1:5 variation of atmospheric pressure, as reported above.

PART II. IMPACT FORCES

§ 1. THEORETICAL ASPECTS OF IMPACT ON LIQUIDS

There are two aspects from which one can regard the impact, both based on the idea of the "added mass" of liquid which the body sets in motion. On the older theory, it is supposed that when a body of mass M enters a liquid there is a sudden reduction of its velocity which is ascribed to an apparent addition to its mass by the mass of liquid m set instantaneously in motion. Thus if V_0 is the velocity just before impact, V_1 just after,

$$MV_0 = (M + m)V_1, \quad \dots\dots(3)$$

or if the volume of liquid is expressed as a fraction μ of the volume of the solid and σ is the specific gravity of the solid relative to the liquid,

$$\sigma(V_0 - V_1) = \mu V_1 \quad \text{and} \quad \mu = \sigma(V_0/V_1 - 1). \quad \dots\dots(4)$$

Theoretical values of m are known for three shapes (cf. Lamb, *Hydrodynamics*, §§ 108, 92 and 71); the disc ($8c^3/3$), the sphere ($2\pi c^3/3$) and the plate of infinite aspect ratio (πc^2 per unit length), where $2c$ is the width. It will be noted that this conception makes the impact force infinite.

In von Kármán's and Wagner's adaptations (1932) of the "added mass" idea, it is assumed that m gradually increases during immersion and is at any instant equal to the mass of liquid contained in a hemisphere or hemi-cylinder erected on the wetted perimeter. Knowing how the latter depends on the depth of immersion and the instantaneous rate of immersion (V), one can calculate the force-time curve for the impact in the form of $\partial(mV)/\partial t$ against t .

§ 2. CALCULATION OF IMPACT FORCES

If an axially symmetrical body hits a liquid at a speed V_0 , the force at any subsequent instant

$$F = -M \frac{\partial V}{\partial t} = \frac{\partial}{\partial t}(mV) = \frac{\partial}{\partial t}(\frac{2}{3}\pi c^3 \rho V) = \frac{2}{3}\pi c^3 \rho \frac{\partial V}{\partial t} + 2\pi c^2 \rho V^2 \frac{\partial c}{\partial h},$$

where h is the depth of immersion and c the radius of the wetted perimeter.

Example 1.—Sphere.

$$h = a - \sqrt{(a^2 - c^2)}, \quad \partial h/\partial c = c/\sqrt{(a^2 - c^2)},$$

whence

$$F = -mF/M + 2\pi c \sqrt{(a^2 - c^2)} \rho V^2,$$

or (with (3))

$$\frac{F}{\rho V_0^2} = \frac{2\pi c \sqrt{(a^2 - c^2)}}{(1 + m/M)^3}. \quad \dots\dots(5)$$

The maximum value of this when M is large compared to m occurs when $c = a/\sqrt{2}$ and is then πc^2 .

Example 2.—Cone of semi-angle γ .

$$\partial h/\partial c = \cot \gamma, \quad F/\rho V_0^2 = 2\pi c^2 \tan \gamma / (1 + m/M)^3.$$

The maximum value of this occurs when $m = \frac{1}{2}M$ and is then $0.86c^2 \tan \gamma$.

It will be noticed that this theory would still make the impact force of a *disc* infinite, since it takes no account of the sideways escape of liquid to form a splash. Wagner modified the theory in this respect and applied it to obtuse-angled prisms, such as the hull of a sea-plane. The theory probably breaks down at the limit of very acute angles since the "displacement" of liquid by the volume of the immersed segment is then comparable with the added mass of the liquid set in motion.

Kreps (1943) introduced the resistance experienced by the body during immersion. The equation of motion then becomes, neglecting buoyancy, surface tension and compressibility, $M\partial V/\partial t = -\partial(mV)/\partial t + Mg - C_D(\pi c^2)(\frac{1}{2}\rho V^2)$.

When the velocity of impact is small, the last term is unimportant but the gravity term is not, *per contra* when the velocity is very large. In the latter case, the measured decelerations will deviate more from the theoretical as C_D is greater, i.e. for blunt-nosed projectiles.

Another method, due to Shiffman and Spencer (1947), of calculating the impact force on a sphere is derived from consideration of the flow past a lens. It leads to force-time curves of the same general type as those calculated from the Kármán-Wagner theory.

§ 3. SMALL-SCALE IMPACT FORCE EXPERIMENTS

Watanabe (1929, 1930, 1934) at the Okechi Research Laboratory used a piezoelectric gauge mounted on bodies of various shapes, viz. cone, cylinder, sphere and disc, which he dropped on to water. Connections led from the quartz to an amplifier and thence to a cathode-ray oscillograph, on which a force-time curve was inscribed as the body hit the water. Shortly after impact the deceleration reached a maximum whose value was proportional to the square of the impact velocity. He claims that the depth of immersion at which the maximum force is reached is proportional to the impact velocity for each shape, but the points on the graph by which this relationship is deduced are very scattered. When the maximum force was plotted against the weight of the falling body it was found that at

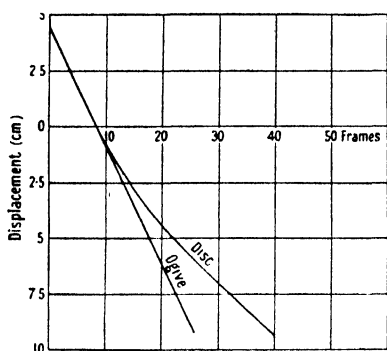


Figure 8. Displacement-time curves on water entry of disc and ogive.

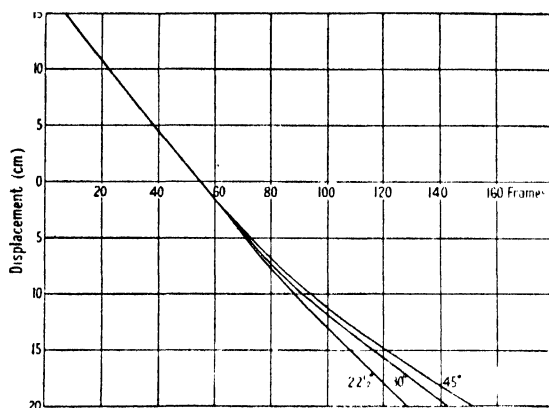


Figure 9. Displacement-time curves on water entry of cones of various semi-angles.

small weights this relationship was one of direct proportionality but subsequently fell away until, when the weight exceeded 3 kg., the impact force was nearly constant, a result which may be connected with the rigidity of the solid body.

At the Glen Fruin Research Station, two techniques have been employed by the author for measuring the change of velocity when the solid hits the water: (a) high-frequency cinematography, which is in most cases best applied to determine the "added mass" as a unique figure for each case, with small projectiles of less than 1 lb. in weight *; (b) an impact-force gauge or decelerometer of high natural frequency applied to a larger hemisphere (20 in. diameter) and capable of delineating the force-time curve accurately.

(a) This series was carried out with a number of hollow ebonite projectiles ($\sigma = 0.7$), 3 in. diameter, dropped freely or fired at speeds up to 125 ft/sec. from a large rubber catapult mounted over a glass-sided tank of water ($2\frac{1}{2}$ feet square in base and 5 ft. deep). Illumination was by searchlight from above and photography by Fastax or Kodak high-speed camera (c. 2000 pictures/sec.). The heads

* Aoki (1926) and Kreps had already used this method, but even for slow entries up to 3 m/sec., such as they used, their picture frequency (60 per sec.) seems inadequate.

of the projectiles were in the shape of a flat disc, a hemisphere, a 1:1 ogive and cones of semi-angles $22\frac{1}{2}^\circ$ and 45° . Spheres of wood and of duralumin were also used in the free-fall experiments. The catapult could produce vertical and oblique entries.

The films were analysed and plotted on graph paper as displacement-time curves. Typical curves are given on figures 8 and 9. In the first method of analysis we ignore—even when we can measure—the rate of change of velocity and draw two straight lines on the displacement-time curves to represent the velocities before and after impact. The main impact is felt when the nose is 1 cm. or more below the water level, depending on the shape of the head.

Table 2 summarizes these measurements; m is expressed both as a fraction of M and as λc^3 , where c is the radius of the cylindrical (or spherical) portion of the head. (Thus the theoretical values of λ are 2.66 for the disc and 2.1 for the hemisphere.)

Table 2

Model head	M (gm.)	Specific gravity σ	β_0 (deg.)	V_0/V_1	m (gm.)	λ_{40}	λ_4
Hemisphere	725	0.73	90	1.2	145	2.7	1.8
			64	1.25	180	3.4	—
			35	1.3	220	4.1	—
Disc	620	0.57	90	1.5	310	6.0	4.0
			64	1.4	250	4.7	—
			35	1.3	185	3.6	—
$22\frac{1}{2}^\circ$ cone	620	0.57	90	1.15	95	1.8	1.5
			64	1.15	95	1.8	—
			35	1.2	125	2.3	—
45° cone	550	0.60	90	1.3	165	3.2	3.0
			64	1.2	110	2.0	—
			35	1.15	85	1.6	—
1:1 ogive	710	0.75	90	1.0	0	0.0	0.0
			64	1.05	35	0.6	—
			35	1.1	70	1.3	—

M =mass of model ; σ =specific gravity ; β_0 =entry angle ; m =added mass ;
 $\lambda_{40} = \lambda$ at 40 m/sec. ; $\lambda_4 = \lambda$ at 4 m/sec.

It will be noted that the derived values of added mass for vertical entry at high speed are greater than those at low. Examination of the photographs showed that this difference was reflected in the amount of water which was able to get away after impact in the splash. The flat disc, of course, produces a large splash and has a correspondingly high value (compared with theory) for λ .

Whether one regards this apparent variation of added mass with speed as significant depends on the attitude one adopts to this method of analysis. It must be again emphasized that "added mass" is rather an academic concept and, as figures 8 and 9 show, difficult to apply in practice, owing to the uncertainty in sighting the straight line which is to give the "velocity just after impact". It is indeed only on the basis of an insignificant drag that one can apply this idea.

It is rather surprising that the ogive enters vertically with inappreciable deceleration until drag begins to be felt.

Considering the results for cones it appears that, within the limits of experimental accuracy, a given cone experiences its impact at the same depth for all entry speeds. The greater the angle of the cone, the nearer to the vertex does this point occur.

(b) The projectile for this series of experiments was a hollow hemisphere of laminated plywood 20 in. in diameter (plate II, figure 10) with the impact gauge bolted rigidly to the inside. A lid prevented entry of water. The tare weight was 24 lb., but lead masses could be bolted to the inside to increase it to 46 lb. Calculation from (5) showed that the peak deceleration could be expected at an immersion of less than half an inch, equivalent to 5 msec. at 8 ft/sec. It was therefore necessary to use an accelerometer of quick response if the peak were to be truly recorded. After some early attempts with purely mechanical gauges, use was made of an electronic instrument developed by the Instrument Department of the Royal Aircraft Establishment from one described by Brookes-Smith and Colls (1939). This consists of a small electrical capacitor, one plate of which is rigidly attached to the structure while the other, in the form of a diaphragm, carries a free load and thus forms the spring + mass element of the accelerometer.

Accelerations produce relative changes in the separations of the plates, and the resulting change in capacitance is used to modulate an oscillatory circuit, which in turn causes movement of the electron spot on the screen of a cathode-ray oscillograph. The motion of the spot is recorded by a camera with continuously moving film.

The gauge had a natural frequency of about 900 c/sec., but filters could be incorporated to cut out the natural oscillations. At a preliminary trial the hemisphere was dropped from a small height three times in succession, the conditions of recording being (a) no filter, (b) filter partially applied, (c) oscillation just damped out. Figure 11 shows the corresponding records, on which the relative maxima of the smoothed curves are (a) 2.6, (b) 2.5, (c) 2.0. In order not to lose a considerable proportion of the true deflection, condition (b) was adopted throughout the remaining trials.

The characteristics of the filter were known. The cut-off was zero at 50 c/sec., and at 100 c/sec. had reached 20% with the full filter applied. It will be noticed in the force-time records (figures 12-14) that the peak was reached in 3.5 to 10 msec., corresponding to quarter-cycles of frequencies 60 to 25 c/sec. Over this range the correction for reduced response by the inclusion of the filter could be neglected.

Calibration was effected by noting the difference between the zero position of the spot on the film with the sphere at rest and that in free fall, set equal to 1g. and confirmed by the addition of known masses to the accelerometer diaphragm. Time marks were given by a 100 c/sec. fork.

The impact was photographed, as before, by Fastax camera.

The experimental records, whether by accelerometer or by camera, are obtained in terms of time and not depth of immersion, as in the theory. In order, then, to compare with theory, velocity-time curves are derived from the photographic records, and these are used to convert the time base of the oscillograph records into depths of immersion. This "correction for drag"—as it in fact amounts to—makes little difference to that value which one would have obtained by assuming the velocity had remained V_0 up to the instant of maximum force. Thereafter drag causes a progressive discrepancy between the actual and the theoretical force-time curves.

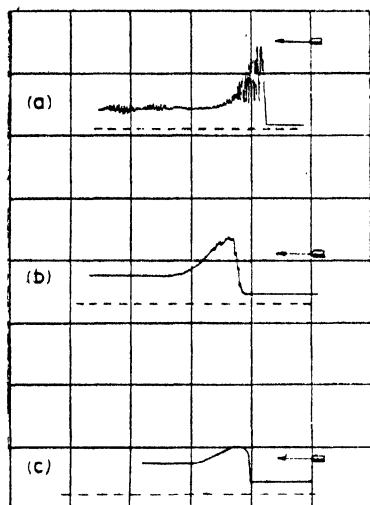


Figure 11. Trial deceleration records under three conditions: no filter (a); partial filter (b); critically damped (c).

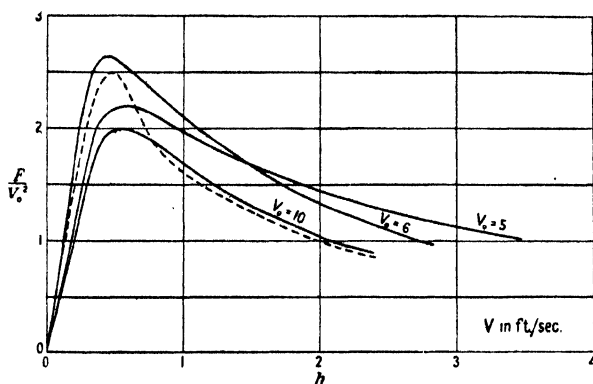


Figure 12. Impact of hemisphere (mass 24 lb.): force for unit entry velocity plotted against depth of immersion.

Typical curves for the 24-lb. mass when dropped from 6 in., 12 in. and 24 in. are shown in figure 12 and for the 46-lb. mass in figure 13. By comparing these with those derived from (5) (which are shown by broken lines) it appears that the von Kármán hypothesis gives good agreement with practice at least up to the instant of maximum impact.

As the ciné-camera was provided with a time base it is possible to derive acceleration-time records. The accuracy of this alternative method depends in the main on the frequency of the pictures in relation to the maximum deceleration. In figure 14 are shown two of the drops as interpreted from the camera records at 2000 frames/sec. To get a film record comparable in accuracy with the gauge record a picture frequency of double this would have been desirable.

The sphere involves a comparatively gentle rise to the maximum force, and with such shapes and a convenient picture frequency the photographic method is

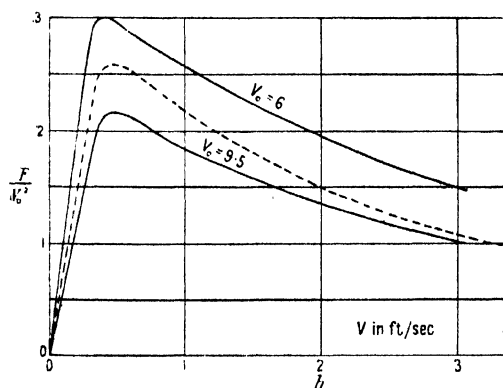


Figure 13. Impact of hemisphere (mass 46 lb.): force for unit entry velocity plotted against depth of immersion.

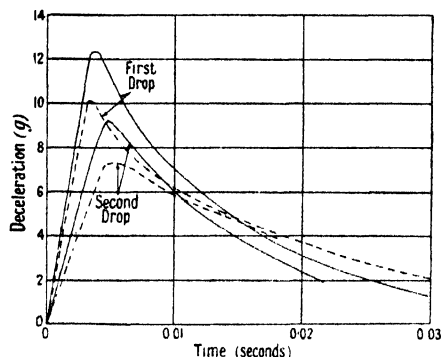


Figure 14. Impact of hemisphere (mass 24 lb.): comparison of camera record (full lines) and decelerometer record (broken lines).

suited to the deduction of impact forces. For sharp impacts the requisite rapidity is at present unattainable. Attempts to analyse the photographs of the impact of a disc in this way failed. At the other extreme of a finely pointed body, as already noted, the deceleration is too small to detect.

§ 4. RICOCHET OF A SPHERE

In order to study the mechanism of ricochet and the form of cavity associated with entry at glancing angles a number of shots were made by firing 3 in. ebonite and duralumin spheres from the catapult, and 1 in. steel spheres from a gun, into the Glen Fruin tank. This permitted a water trajectory of about 20 ft. which could be photographed in two sections by ciné-cameras running at a speed of 200 pictures per second. Entry speeds ranged up to 360 ft/sec. The trajectories and displacement-time curves were plotted, and entry and exit angles (β_0, β_1 radians) and velocities (V_0, V_1), if ricochet occurred, were noted and the mean drag and lift coefficients between make-and-break contact over the path s were calculated using the formulae

$$C_D = (3.05\sigma d/s) \log(V_0/V_1), \quad \dots\dots(7)$$

$$C_D = 1.33\sigma d(\beta_0 + \beta_1)/s. \quad \dots\dots(8)$$

In another experiment the 20 in. wooden hemisphere, with its electrical accelerometer attached, was allowed to "glide" on to the water in order that the

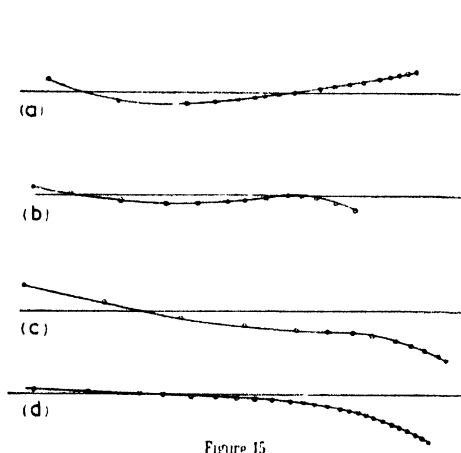


Figure 15

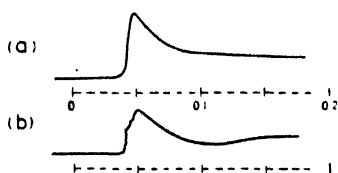


Figure 17.

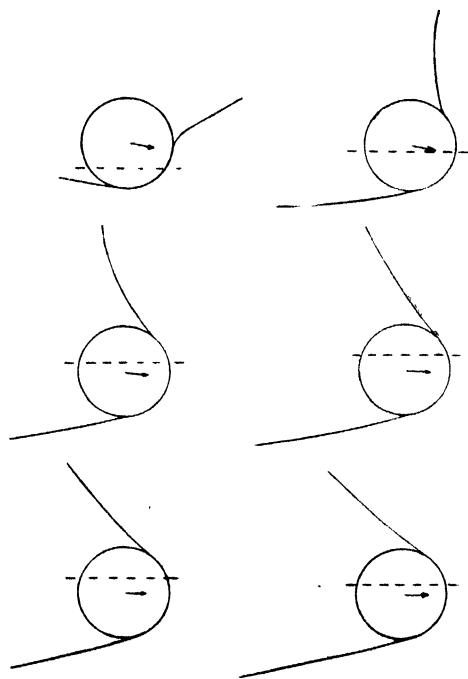


Figure 16.

Figure 15. Types of trajectory at glancing entry, the circles showing successive positions of the solid.

Figure 16. Water-surface shapes round spheres at glancing entry, the dotted lines indicating water level before being disturbed.

Figure 17. Decelerometer records, entry at 90° (a), at 28° (b) to horizon.

vertical component of force on a sphere entering obliquely could be measured. This entry was also photographed.

From all the photographs the shape of the splash and the wetted areas at different epochs were noted.

One may distinguish four types of trajectory as the entry angle is increased. Examples of these are shown in figure 15: (a) definite ricochet with angle of exit somewhat less than angle of entry; (b) break-surface often followed shortly by re-entry; (c) flattening-out and continuing on a straight path for a certain distance; (d) continuing straight ahead, then diving. (In the figure the dots represent positions on successive frames of the film, and so enable the deceleration to be judged.) The data for those shots which were capable of analysis are set out in table 3.

Table 3

No.	Sphere, <i>d</i> and σ	Velocity in path, (ft/sec.)		Angle to water surface (deg.)		Path length (in.)	Max. depth (diam.)	Mean C_D C_L	
	Dural	V_0	V_1	β_0	β_1	S			
1	3'', 2.7	100	50	7½	4½	50	0.5	0.15	0.045
2	"	120	33	10	6	60	0.5	0.23	0.055
3	"	100	53	9½	4	35	0.75	0.20	0.076
4	"	130	65	7½	6	55	0.75	0.14	0.042
6	"	100	65	7	5	42	—	0.11	0.052
15	"	147	52	9	10½	84	0.75	0.13	0.048
16	"	120	30	5½	4½	120	1.0	0.12	0.015
17	"	100	25	6	9	90	1.0	0.17	0.028
23	"	135	67*	8½	(0)*	60*	3.5	0.12	0.027
	Steel								
32	1'', 7.8	205	55*	7	(0)*	72*	2.0	0.19	0.013
36	"	210	140	8	4	66*	1.0	0.06	0.033
37	"	310	140	7½	4	60	0.5	0.14	0.035
38	"	330	140	6½	5	30	0.75	0.30	0.062
39	"	150	50*	11½	—	54*	—	0.21	—
40	"	270	80*	6½	(0)*	100*	1.8	0.13	0.012
42	"	300	65*	6	—	72*	1.5	0.22	—
43	"	360	55*	5	(0)*	76*	1.5	0.25	0.012
	Ebonite								
29	3'', 1.1	135	35	12½	10	160	1.25	0.04	0.011
47	"	155	58	17	11	55	1.25	0.08	0.036
48	"	155	37	14	14	48	1.5	0.13	0.043
61	"	142	56	11½	3½	42	0.7	0.10	0.027
70	"	135	16	6	6	54	0.8	0.17	0.018
67	"	142	—	12	—	—	—	—	—
69	"	140	—	12½	—	—	—	—	—

* Notes: Shot 23 flattened in 60''; 32 flattened in 72''; 39 straightened out; 40 flattened in 100''; 42 straightened out; 43 flattened in 76''; 67 and 69 broke surface and re-entered.

Besides the shots which were photographed and completely analysed, some other visual observations of shots at various entry angles were made, merely to see whether the projectile broke surface or remained submerged. By grouping these observations, the following values of the critical angles for ricochet (which did not vary significantly with speed) were deduced: steel ($\sigma=7.8$) 6°, dural ($\sigma=2.7$) 9°, ebonite ($\sigma=1.1$) 15°.

These results agree with some early observations of de Jonquières (1883) on steel spheres in that the product of the critical angle and the square root of the specific gravity is approximately constant.

From the photographs a series of tracings of the shape of the water surface at equal steps along the water path have been made. Figure 16 shows a typical set. It will be noted that the wetted perimeter forms a circle whose plane is inclined at less than 90° to the direction of motion, so that the reaction of the water has a component producing the lift. The water is able to do this provided the cavity does not reach its ultimate "submerged" shape, with the plane of wetting normal to the trajectory, before the sphere loses speed.

One might try to apply Newton's hypothesis to predict the resistance coefficients in the cavity stage by integrating $\cos^2 \phi \cdot \delta S$ over the wetted surface (δS being an element of the surface whose normal makes an angle ϕ to the trajectory) and resolving the resistance into components parallel and perpendicular to the direction of motion, but, as one would expect, such deduced values of C_D and C_L are much higher than the mean values calculated from (7) and (8); (0.45 and 0.045 in place of 0.12 and 0.027 respectively). Naturally, the drag coefficients deduced from the trajectory in oblique entry fall below that (0.34) appropriate to vertical entry of the sphere.

Figure 17(b) shows a decelerometer record for the entry of the 20-in. wooden hemisphere at 28° to the water surface with a vertical component of velocity (V_0) of 6 ft/sec. The maximum values of F/V_0^2 in the vertical and inclined entries were 2.25 and 1.5 respectively. While the rise to the peak is less sharp in the glide, the deceleration falls off at about the same rate in the two cases. Apparently forward momentum produces a "cushioning" of the water and a lifting force even in the short time that elapses before the maximum impact (in a vertical drop) is reached.

§ 5. FULL-SCALE APPLICATIONS OF IMPACT EXPERIMENTS

Fundamental experimental work on the impact of solids on water has two principal applications: (i) in ballistics, to the passage of projectiles from air into water; (ii) in aeronautics, to the landing of seaplanes.

The work described herein had principally the first application in view. Although the speeds and sizes of projectiles in practice are mostly greater than those employed in these experiments, the considerable range of these variables covered has indicated, for the sphere, that a model to full-scale conversion may be applied on a Froude number basis, where air-to-water ballistics is concerned. The underwater trajectory of longer projectiles which, in the cavity stage, involves moments set up by tail forces is another problem to which much research has been devoted.

The slow motion of the large hemisphere into water is more concerned with the second application. The theories of von Kármán and Wagner were in fact devised with this object in view, the keel of the boat being idealized into a shallow wedge for this purpose. Although it should be possible, using high-speed ciné-photography, to study impact forces on full-sized seaplanes dropped on to water, as far as the author knows this has never been done. The seaplane designer is more concerned with the local stresses set up at the touch-down, and these have been studied up to the present by observing pressure gauges let into various places on the

keel and flush with its surface. Two reports by Jones *et al.* (1934, 1937) describe measurements of this type using mechanical gauges. Recently these have been repeated using the electronic gauge as applied to the hemisphere in the present work.

ACKNOWLEDGMENTS

The author thanks the Director of the Royal Aircraft Establishment for permission to publish this work. He also acknowledges his indebtedness to a number of colleagues of the R.A.E. for their willing assistance, in particular to Mr. D. D. Hardy, laboratory assistant at Farnborough, and to Mr. F. P. Mayo, technical assistant at Glen Fruin.

REFERENCES

- AOKI, T., 1922, *J. Coll. of Engng., Tokyo*, **11**, 10 ; 1926, *Ibid.*, **16**, 4, 17, 4.
 BAUER, W., 1926, *Ann. Phys., Lpz.*, **82**, 1014.
 BELL, G. E., 1924, *Phil. Mag.*, **48**, 753.
 BROOKES-SMITH, C. H. W., and COLLS, J. A., 1939, *J. Sci. Instrum.*, **16**, 361.
 VON KÁRMÁN, TH., 1930, *N.A.C.A. Tech. Note* 32.
 KREPS, R. L., 1943, *N.A.C.A. Tech. Mem.* 1046.
 JONES, E. T., and DAVIES, W. H., 1934, *Aero. Res. Commn. R. and M.* 1638.
 JONES, E. T., DOUGLAS, G., STAFFORD, C. E., and CUSHING, R. K., 1937, *Aero. Res. Commn. R. and M.* 1807.
 DE JONQUIÈRES, E., 1883, *C.R. Acad. Sci., Paris*, **97**, 1278.
 LEHMANN, K. O., 1934, *Ann. Phys., Lpz.*, **21**, 533.
 McMILLEN, J. H., 1945, *Phys. Rev.*, **68**, 198.
 MALLOCK, A., 1918, *Proc. Roy. Soc. A*, **95**, 138.
 PABST, W., 1931, *Z. Flugtechnik*, **22**, 193.
 RAMSAUER, C. (and DOBKE, G.), 1927, *Ann. Phys., Lpz.*, **84**, 697.
 SCHIFFMAN, M., and SPENCER, D. C., 1947, *Quart. Appl. Maths.*, **5**, 270.
 TREY, F., 1933, *Phys. Z.*, **34**, 494.
 WAGNER, H., 1932, *Z. angew. Math. Mech.*, **12**, 193.
 WATANABE, S., 1929, *Sci. Pap. Inst. Phys. Chem. Res., Tokyo*, **12**, 251 ; 1930, *Ibid.*, **14**, 153 ; 1934, *Ibid.*, **23**, 118, 202, 249.
 WORTHINGTON, A. M., 1877, *Proc. Roy. Soc.*, **25**, 261 ; 1897, *Philos. Trans. A*, **189**, 137.

A Correction in Viscometry due to Varying Hydrostatic Head

By J. E. CAFFYN *

The National Institute for Research in Dairying, University of Reading

* Now Lecturer in Physics in The Durham Colleges, The University of Durham

ABSTRACT. The varying hydrostatic head in viscometers gives rise to a correction to the viscosity if the latter is calculated from the applied pressure only. This has been worked out in general and the special cases of viscometers with cylindrical and biconical bulbs and cylindrical bulbs with conical ends have been considered in detail.

§ 1. INTRODUCTION

THE author was led to consider this problem when designing a viscometer for the determination of the viscosity temperature coefficient of milk. v. Mises (1911) gives a correction for variable head for the Engler viscometer which takes account of the kinetic energy of the outflowing liquid.

Simeon (1914) calculates a correction for the Ostwald type viscometer involving the logarithm of the initial and final liquid head driving the fluid through the capillary of the viscometer. Bingham, Schlesinger and Coleman (1916) give a similar treatment for the simple case of a capillary tube connected to two cylindrical bulbs and extend it to the case where each bulb is made up of two cylinders of different radii. Lidstone (1922 a, b, c) deduces a correction for the case considered by Bingham *et al.*, but takes into account the kinetic energy of the liquid. This leads to a formula for the viscosity which can only be evaluated by succession approximation. Bridgman (1927) deduces a similar result, but points out that it is only necessary to use the correction to the pressure due to variable head, and in so far as the latter affects the kinetic energy correction, it is a correction to a correction and can be neglected.

Barr (1924), in discussing Lidstone's viscometer, gives a correction to the pressure arising from variable head for the case of spherical reservoirs without giving the derivation.

Barr (1935) has also given equations for the variable head correction in viscometry and the expansion of the logarithmic correction factors for certain cases. He considered the sphere discharging into air, a pair of opposed cones, a biconical bulb discharging into a cylinder and a cylindrical bulb with conical ends discharging into a cylinder for special values of the parameters.

It appears that no attempt has been made to consider the general case of variable head. It is proposed, therefore, to investigate the case of a viscometer whose bulbs are volumes of revolution. The kinetic energy correction will be neglected in this treatment.

§ 2. GENERAL CASE

Consider a viscometer shown diagrammatically in figure 1. Let the bulbs A and A' be connected by a capillary tube of radius R cm. and length L cm. Let a liquid of viscosity η and density ρ be driven from A to A' by an applied pressure p_0 . The liquid level at any instant in A is h cm. from an arbitrary

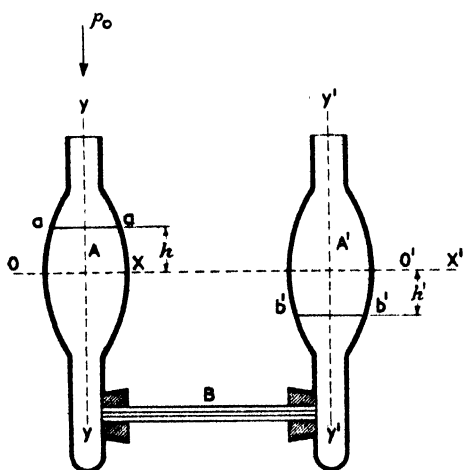


Figure 1.
aa, b'b': initial liquid level.

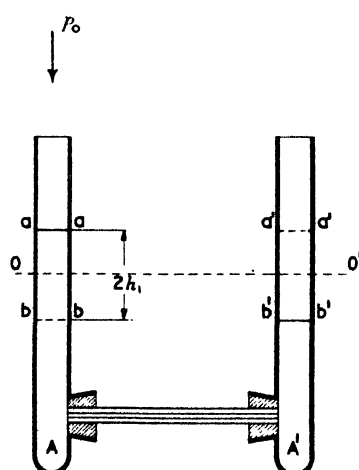


Figure 2.
aa, b'b': initial liquid level.
bb, a'a': final liquid level.

level OO' and the corresponding height of the liquid level in A' is h' . Let the viscometer bulbs A and A' be volumes of revolution generated by rotating the curves $x=f(h)$ and $x'=f'(h')$ about the axes YY and $Y'Y'$ respectively. $f(h)$ and $f'(h')$ need not necessarily be the same functions of h and h' .

Then if we apply Poiseuille's law in its differentiated form to the liquid flowing through the capillary tube B ,

$$dV(h)/dt = \pi R^4 p(h) / 8\eta L ; \quad \dots\dots(1)$$

$V(h)$, the volume of liquid flowing through B , and $p(h)$, the pressure, driving the liquid through B , are both functions of h .

If the liquid level in the bulb A falls dh in time dt , then

$$dV(h) = \pi [f(h)]^2 dh. \quad \dots\dots(2)$$

Therefore, writing (1) in the form

$$(dV(h)/dh) \cdot (dh/dt) = \pi R^4 p(h) / 8\eta L, \quad \dots\dots(3)$$

and substituting from (2), we have

$$\pi [f(h)]^2 (dh/dt) = \pi R^4 p(h) / 8\eta L. \quad \dots\dots(4)$$

The function of $p(h)$ is given in terms of h and h' by

$$p(h) = p_0 + g\rho(h - h'). \quad \dots\dots(5)$$

Substituting from (5) in (4), rearranging and integrating from $h = h_i$, the initial liquid level in A , to $h = h_f$, the final liquid level in A , we have

$$\int_{h_i}^{h_f} \frac{\pi [f(h)]^2 dh}{h_i p_0 + g\rho(h - h')} = \int_0^t \frac{\pi R^4 dt}{8\eta L}. \quad \dots\dots(6)$$

To solve (6) explicitly for the viscosity η , we must specify $f(h)$ and know h' in terms of h . The latter can be determined from the fact that the volume of fluid leaving A must enter A' . Hence

$$\pi \int_{h_0}^h [f(h)]^2 dh = \pi \int_{h'}^{h'_0} [f'(h')]^2 dh', \quad \dots\dots(7)$$

where h_0 and h'_0 are convenient corresponding liquid levels in A and A' . The shape of the viscometer will determine the functions $f(h)$ and $f'(h')$. Then with the specification of the initial and final liquid levels in the bulbs A and A' , the equations (6) and (7) give the solution to the problem.

§ 3. A VISCOMETER WITH CYLINDRICAL BULBS

A viscometer of this type is shown diagrammatically in figure 2. The bulbs have the same radius r cm. Since the initial and final liquid levels are equally displaced about OO' , equation (7) gives on integration $h = -h'$. Therefore equation (6) becomes

$$\pi r^2 \int_{-h_i}^{+h_i} \frac{dh}{p_0 + 2g\rho h} = \frac{\pi R^4}{8\eta L} \int_0^t dt, \quad \dots\dots(8)$$

which on integration gives

$$\eta = \frac{\pi R^4 t p_1}{8LV \ln(p_0 + p_1)/(p_0 - p_1)}, \quad \dots\dots(9)$$

where p_1 is the hydrostatic pressure due to a column of the liquid $2h_1$ cm. high and V is the total volume of liquid passing through the capillary.

Expanding the logarithm in (9), we have

$$\eta[1 + c_1^2/3 + c_1^4/5 \dots] = \pi R^4 p_0 / 8LV, \quad \dots\dots(10)$$

where $c_1 = p_1/p_0$. This shows the effective correction to the applied pressure due to the variable hydrostatic head. This result expressed as in equation (9) is the same as Bingham's. It is not difficult to modify this treatment for the case where the initial liquid level "aa" in A is not at the same horizontal level as the final liquid level "a'a'" in A'.

§ 4. A VISCOMETER WITH BICONICAL BULBS

A viscometer of this type is shown diagrammatically in figure 3. The viscometer is filled so that the liquid level is bb and b'b' in the bulbs A and A'

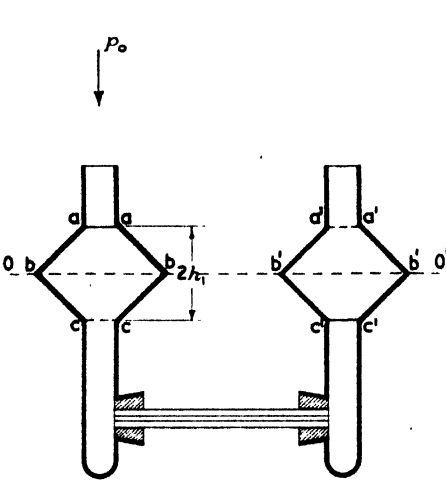


Figure 3.
aa, c'c': initial liquid level.
cc, a'a': final liquid level.

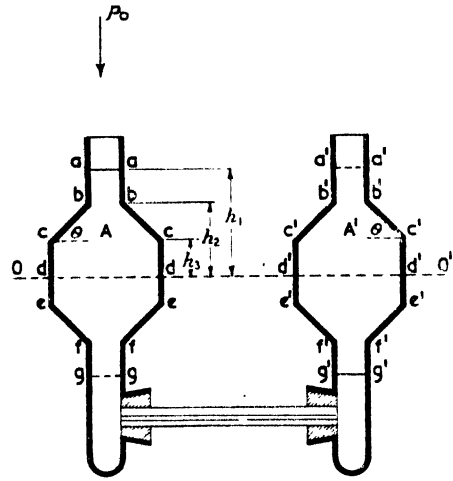


Figure 4.
aa, g'g': initial liquid level.
gg, a'a': final liquid level.

respectively, the bulbs being identical in all respects. The function $f(h) = f'(h')$, takes slightly different forms for positive h and negative h , and is given by

$$f(h) = r_2 - h \tan \theta, \quad h \text{ positive}; \quad f(h) = r_2 + h \tan \theta, \quad h \text{ negative},$$

where r_2 is the radius of the bulb at bb and θ is the angle cb makes with the horizontal. To integrate (6) we must divide up the range of integration, and we get

$$\pi \int_0^{+h_1} \frac{(r_2 - h \tan \theta)^2 dh}{p_0 + g\rho(h - h')} + \pi \int_{-h_1}^0 \frac{(r_2 + h \tan \theta)^2 dh}{p_0 + g\rho(h - h')} = \frac{\pi R^4}{8\eta L} \int_0^t dt. \quad \dots\dots(11)$$

Since the bulbs are identical, equation (7) will give $h = -h'$ for all values of h . Therefore equation (11) becomes

$$\pi \int_0^{+h_1} \frac{(r_2 - h \tan \theta)^2 dh}{p_0 + 2g\rho h} + \pi \int_{-h_1}^0 \frac{(r_2 + h \tan \theta)^2 dh}{p_0 + 2g\rho h} = \frac{\pi R^4}{8\eta L} \int_0^t dt.$$

On carrying out the integrations and making some algebraic simplification, we have

$$\left[\frac{\pi r_2^2}{2g\rho} + \frac{\pi p_0^2(r_2 - r_1)^2}{8g^3\rho^3h_1^2} \right] \ln \frac{p_0 + 2g\rho h_1}{p_0 - 2g\rho h_1} + \frac{\pi r_2(r_2 - r_1)p_0}{2g^2\rho^2h_1} \ln \frac{p_0^2 - 4g^2\rho^2h_1^2}{p_0^2} + \frac{\pi p_0(r_2 - r_1)^2}{2g^2\rho^2h_1} = \frac{\pi R^4 t}{8\eta L},$$

where r_1 is radius of the bulb at aa, etc.

Treating this in the same way as previously, we have

$$\frac{\pi R^4 p_0 t}{8\eta L} = \eta \left[1 + \frac{c_1^2}{10} \left(\frac{1 + 3w + 6w^2}{1 + w + w^2} \right) + \text{terms in } c_1 \text{ of higher powers than the second} \right], \dots (12)$$

where w is ratio of the radius of the bulb at "aa" to the radius at "bb" and c_1 has its usual meaning. As in the previous case, suitable modification of the treatment will give the solution for the case when the centres of the bulbs "bb" and "b'b'" do not lie in the same horizontal plane.

§ 5. A VISCOMETER CONSISTING OF CYLINDRICAL BULBS WITH CONICAL ENDS

This shape of bulb is shown diagrammatically in figure 4 and is approximately similar to many industrial viscometer bulbs. As previously, we will consider the case of identical bulbs, i.e. $f(h) = f'(h')$, and hence $h = -h'$ for all values of h . The function $f(h)$ is defined in the ranges as follows:

$$f(h) = r_1, \quad h_2 < h < h_1; \quad f(h) = r_2, \quad -h_3 < h < h_3; \quad f(h) = r_1, \quad -h_2 < h < -h_1$$

$$f(h) = r_1 + (h_2 - h) \tan \theta, \quad h_3 < h < h_2; \quad f(h) = r_1 + (h_2 + h) \tan \theta, \quad -h_3 < h < -h_2,$$

where r_1 and r_2 are the radii of the bulbs at "aa" and "dd" respectively. Splitting up the range of integration, substituting the appropriate values of $f(h)$ in equation (6) and integrating, we have

$$\begin{aligned} \frac{\pi R^4 t}{8\eta L} = & \frac{\pi r_1^2}{2g\rho} \ln \left[\frac{p_0 + p_1}{p_0 + p_2} \cdot \frac{p_0 - p_2}{p_0 - p_1} \right] + \frac{\pi r_2^2}{2g\rho} \ln \frac{p_0 + p_3}{p_0 - p_3} \\ & + \frac{\pi(r_1 + h_2 \tan \theta)^2}{2g\rho} \ln \left[\frac{p_0 + p_2}{p_0 + p_3} \cdot \frac{p_0 - p_3}{p_0 - p_2} \right] \\ & + \frac{\pi p_0 \tan \theta (r_1 + h_2 \tan \theta)}{2g^2\rho^2} \ln \left[\frac{p_0 + p_2}{p_0 + p_3} \cdot \frac{p_0 - p_2}{p_0 - p_3} \right] \\ & + \frac{p_0^2 \pi \tan^2 \theta}{8g^3\rho^3} \ln \left[\frac{p_0 + p_2}{p_0 + p_3} \cdot \frac{p_0 - p_3}{p_0 - p_2} \right] - \frac{\pi p_0 \tan^2 \theta (h_2 - h_3)}{2g^2\rho^2}, \dots (13) \end{aligned}$$

where p_1 , p_2 and p_3 are the pressures due to liquid columns $2h_1$, $2h_2$ and $2h_3$ cm. high respectively.

If we expand the logarithms in equation (13) and rearrange, we can write equation (13) in the form

$$\eta \left[1 + \frac{V_1}{V} (c_1^2 + c_1 c_2 + c_2^2) + \frac{V_3}{3V} (c_3^2) + \frac{V_2}{10V} \left\{ \frac{(6c_3^2 + 3c_3 c_2 + c_2^2) + w(3c_3^2 + 4c_3 c_2 + 3c_2^2) + w^2(6c_2^2 + 3c_2 c_3 + c_3^2)}{1 + w + w^2} \right\} + \text{higher terms} \right] = \frac{\pi R^4 p_0 t}{8VL}, \quad \dots (14)$$

where V_1 is the volume "aabb" + "ffgg", V_2 , the volume "bbcc" + "ceff", V_3 , the volume "ccee", $c_1 = p_1/p_0$, $c_2 = p_2/p_0$, $c_3 = p_3/p_0$; $V = V_1 + V_2 + V_3$, and the other symbols have the same meaning as previously. Equations (14) and (12) reduce to equation (10) on substituting $w=1$, V_1 in terms of h_1 , etc.; this provides a convenient check on the algebra.

§ 6. DISCUSSION

In the table the correction to the viscosity, as calculated from the applied pressure only, is tabulated against c_1 for the three types of viscometer considered above with suitable arbitrary values of the constants. The value of c_1 must always be less than one, otherwise the logarithmic expansion would not be justified. This means physically that the applied pressure must be greater than the maximum hydrostatic head, which is necessary if the experiment is to end in a finite time.

Correction factors to η

c_1		0.200	0.100	0.050	0.033	0.025
Correction factor for viscometer shown in	figure 2	1.0133	1.0033	1.0008	1.0004	1.0002
	figure 3	1.0020	1.0005	1.0001	<1.0000	<1.0000
	figure 4	1.0052	1.0013	1.0003	1.0002	1.0001

Values for viscometer constants used in determining the corrections above: $w=2/5$; $c_2=11c_1/12$; $c_3=2c_1/3$; $V_1/V=1/30$; $V_2/V=1/5$; $V_3/V=1/10$ and $\theta=45^\circ$.

ACKNOWLEDGMENT

The work described above was part of a research programme carried out at the National Institute for Research in Dairying under a grant from the Agricultural Research Council.

REFERENCES

- BARR, G., 1924, *J. Soc. Chem. Ind.*, **43**, 29; 1935, *J. Chem. Soc.*, **17**, 94.
 BINGHAM, E. C., SCHLESINGER, H. I., and COLEMAN, A. B., 1916, *J. Amer. Chem. Soc.*, **38**, 27.
 BRIDGEMAN, P. W., 1927, *Proc. Amer. Acad. Arts Sci.*, **62**, No. 8, 187.
 LIDSTONE, F. M., 1922 a, *Phil. Mag.*, **43**, 354; 1922 b, *Ibid.*, **43**, 1024; 1922 c, *Ibid.*, **44**, 953.
 v. MISES, F., 1911, *Phys. Z.*, **12**, 812.
 SIMEON, F., 1914, *Phil. Mag.*, **27**, 38.

The Radial Variation of the Earth's Magnetic Field

By S. K. RUNCORN

The Physical Laboratories, The University of Manchester;

With an Appendix by S. CHAPMAN

The Queen's College, Oxford

Communicated by P. M. S. Blackett; MS. received 30 April, 1948

ABSTRACT. It is shown that Core and Distributed theories of the Earth's Magnetic Field predict different variations of field intensity with depth. Whereas on core theories both the horizontal and vertical intensities increase with depth according to an inverse cube law, on a distributed theory such as the one recently put forward by Blackett we find, with reasonable assumptions, that while the vertical intensity should increase for small depths as an inverse cube law, the horizontal intensity should decrease. The difference between the theories is even more marked deep inside the Earth. It is also suggested that if the variation with depth experiments gives a positive result in favour of distributed theories, then experiments on the magnetic effect of gorges in the Earth's surface might throw further light on the mechanism of the phenomena.

§ 1. INTRODUCTION AND SUMMARY

THE theories of the main magnetic field of the Earth may be divided roughly into two groups.

(i) *Core theories*—those which attribute the field to processes taking place in or properties possessed by the central core of the earth and

(ii) *Distributed theories*—those which attribute the field to a process or property associated with the whole volume of the Earth. Such a distributed theory is that due to Blackett (1947) who, by noting the proportionality of magnetic moment and angular momentum for the Earth, the Sun and a star (78 Virginis), has lately revived an old idea of Schuster and H. A. Wilson and tentatively postulates the Earth's main magnetic field to be an example of a universal property of rotating matter. Bullard pointed out that it might be possible to decide between the core theories and a distributed theory as their laws of variation of field with depth would be different. This paper develops this idea and compares the law of variation with depth for the two groups of theories.

§ 2. CORE THEORIES OF THE EARTH'S MAIN MAGNETIC FIELD

The theories which have been put forward at various times to explain the Earth's magnetism have been summarized by Chapman and Bartels (1940), by J. A. Fleming and others (1940), and more recently by Blackett (1947).

Seismological evidence shows that the Earth consists of a solid crust and a liquid core of diameter about 2000 miles. The mean density of the liquid core (12) is the density that iron and nickel would have at the pressures obtaining in the core. Thus it is thought that the core might consist of iron or iron-nickel. As the effect of very high pressure on the Curie point of iron is not known the hypothesis that the Earth's core is permanently magnetized to an intensity of about 0.6 gauss cannot be excluded completely. (If the whole of the Sun was uniformly magnetized the intensity of magnetization required to explain its observed field would have to be about 12 gauss). Nippoldt and Bartels (see Fleming 1940) have suggested that the entire field has its origin in ferromagnetism

of the outer crust down to the point where the temperature has increased to the Curie point for iron. This involves attributing an average intensity of magnetization of the crust of between 1 and 6 gauss. If this occurs it must lie below the depths normally accessible to observation and hence the theory would give the same variation with depth as a core theory.

The field of the Earth has also been attributed to electric conduction currents. However, Lamb showed that the decay of such currents in the earth would be far too rapid to attribute them to a primeval field. But there are no known mechanical motions or chemical reactions in the outer crust of the Earth which would maintain such currents, hence Elsasser (1939, 1941) and Frenkel (1945) have postulated processes in the central core of the Earth resulting from its fluid state, e.g. convective motions leading to thermoelectric E.M.F.s which might maintain currents sufficient to explain the Earth's magnetic field.

§ 3. THE RADIAL VARIATION FOR CORE THEORIES

If the magnetic field originated in the central core of the earth or below the deepest point at which our measurements can reach, as in the theories listed in § 2, the field would vary inversely with the cube of the distance from the centre of the Earth.

If R is the radius of the Earth, H^{θ} is the horizontal intensity, H^r is the vertical intensity, suffix d refers to measurements at depth d and suffix 0 to measurements at the surface, we have at co-latitude θ

$$H_d^r = H_0^r(1 + 3d/R); \quad H_d^{\theta} = H_0^{\theta}(1 + 3d/R) \text{ for small } d/R.$$

Thus in Great Britain we would expect the horizontal intensity to increase by 12γ ($1\gamma = 10^{-5}$ gauss) and the vertical intensity by 26γ at a depth of 4000 ft. (which is the depth of the deepest mine in this country). All measurements at depth are assumed to be made in filamentous cavities parallel to the component of the field being measured, to avoid the effects due to the susceptibility of the surrounding rocks, which usually lies between 10^{-6} and 10^{-3} gauss.

§ 4. THE DISTRIBUTED THEORIES OF THE EARTH'S MAIN MAGNETIC FIELD

The simplest distributed theory is to assume that the Earth possesses a uniform negative volume charge of 4×10^{-4} E.S.U./cm³ or a surface charge of 5.1×10^4 E.S.U./cm². However, these involve enormously strong electric fields at the surface of the Earth of the order of 10^8 volt/cm. Moreover, as Schuster (1912) points out, such a theory would give a field radically different from that of a dipole for an observer on the Earth's surface. Sutherland (1900) overcame both of these objections by suggesting that the Earth might possess a negative surface charge and a positive volume charge sufficient to neutralize the external electric field. Such a theory involves an electric field of 10^8 volt/cm. inside the Earth. Sutherland (1903) then proceeded to show that if the charge separation took place at the atomic or molecular level rather than macroscopically it would be necessary to assume that the atomic charges separate by amounts of the order of 10^{-8} cm. Schlomka (1933) and Haalck (1937, 1938) work out theories on these lines, making assumptions as to the cause of the charge separation.

Blackett (1947), on the basis of the most recent measurements of the field of the Sun and 78 Virginis, tentatively puts forward the view that all massive

rotating bodies have associated with them a magnetic field which he assumes in the case of a spherical body to be, externally to the body, the field of a dipole situated at the centre directed along the axis of rotation of moment

$$P = -\beta U G^{\frac{1}{2}} / 2c, \quad \dots\dots(1)$$

where G is the gravitational constant, U the angular momentum, c the velocity of light, and β a constant of order unity. Such a theory implies that each particle of matter in the body makes a contribution to the magnetic field. And Blackett argues that owing to the failure of all classical attempts to provide an explanation of the formula, the law (1) may imply some new fundamental relation between electromagnetism and gravitation.

In addition to Blackett's hypothesis, a complete explanation of geomagnetism would have to postulate certain secondary mechanisms which would give rise to declination and secular variation in addition to the field components of higher harmonic orders (Elsasser 1941, Bullard, unpublished). In this paper such secondary effects will be ignored.

§ 5. RADIAL VARIATION FOR DISTRIBUTED THEORIES

Since the deepest mines in the world are 6000 ft. deep, bore holes reach 14000 ft. in depth and the greatest depths in the oceans are of the order of 34000 ft., the determinations of the variation of the Earth's field with depth may only be carried out at depths small compared to the radius of the Earth. Let the magnetic moment of the Earth be P and let P_d be the magnetic moment of the surface shell S of thickness d inside which measurements are made. It will be assumed that the magnetic effects of concentric shells of the Earth are additive for distributed theories, and hence as the main magnetic field is that of a dipole the field of the shell S alone would be that of a dipole outside the sphere.

The potential of the isolated rotating shell for $r > R$ would be

$$\begin{aligned} \phi_0 &= P_d \cos \theta / r^2, \text{ where } \theta \text{ is the co-latitude angle.} \\ &= (P_d / r^2) P_1(\cos \theta), \text{ where } P_1 \text{ is the first zonal harmonic. } \dots\dots(2) \end{aligned}$$

Now the field within and without the shell may be represented by scalar potentials expanded in the form of series of zonal harmonic terms in their rôle as solutions of Laplace's equation :

$$\begin{aligned} \phi_1 &= I_n(r/R) P_n(\cos \theta) \text{ for } r < R, \\ \phi_0 &= E_n(R/r)^2 P_n(\cos \theta) \text{ for } r > R, \end{aligned}$$

where E_n and I_n are coefficients independent of r and θ . Hence $E_n = P_d/R^2$ for $n=1$, and $E_n=0$ for $n>1$.

To determine the coefficients I_n it is necessary to insert a boundary condition. On classical magnetostatic theory there are two alternatives, the shell is to be regarded (A) as a virtual surface distribution of current in which case the radial magnetic induction passes continuously through the shell (Stratton 1941, §4.12) or (B) as a surface distribution of magnetic moment. In the absence of an explanation of formula (1) it is finally necessary to rely on the experimental determination of the variation of the components of the Earth's magnetic field with depth to determine the nature of the boundary condition. In this paper we will assume that (A) is the right boundary condition to use. Thus

$$I_1 P_1(\cos \theta) / R = -2(P_d / R) P_1(\cos \theta) \text{ and } I_n = 0 \text{ for } n > 1.$$

Thus we obtain the scalar potential inside a thin shell of matter of moment P_d

$$\phi_i = -2(P_d/R)rP_1(\cos\theta). \quad \dots\dots(3)$$

Now the field intensities at the surface of the sphere which would remain if the outer shells were removed are

$$H^r = + \frac{2(P - P_d)}{(R - d)^3} \cos\theta \quad \text{and} \quad H^\theta = + \frac{(P - P_d)}{(R - d)^3} \sin\theta.$$

Now for small values of d/R the radial component of the field is continuous through the shell, hence the vertical field at depth d is $H_d^r = H_0^r(1 + 3d/R)$ where H_d and H_0 denote the field intensities at depth d and the surface of the Earth respectively.

Just inside the shell S the horizontal field H^θ due to the shell is given by $H^\theta = [(1/r)(\partial\phi_i/\partial\theta)]_{r=R} = 2P_d/R^3 \sin\theta$ and just outside the shell by

$$H^\theta = [(1/r)(\partial\phi_0/\partial\theta)]_{r=R} = -(P_d/R^3) \sin\theta.$$

Summing the field components we find that the horizontal component at depth d below the surface of the Earth is $H_d^\theta = H_0^\theta(1 + 3d/R - 3P_d/P)$. Measurements of the variation of the horizontal component with depth gives a direct measurement of the magnetic moment of the shell S.

§ 6. RADIAL VARIATION FOR BLACKETT'S FUNDAMENTAL THEORY

By the use of law (1) and boundary condition (A) the variation of the magnetic field throughout the Earth may be calculated.

We begin by assuming that the contribution (dP) made to the total dipole by a shell of matter at radius a in the sphere may be obtained by differentiating the expression for P

$$dP = \beta \frac{G^{\frac{1}{2}}}{2c} dU = \frac{4\pi}{3} \beta \frac{G^{\frac{1}{2}}}{c} \rho a^4 \omega da, \quad \dots\dots(4)$$

where da is the radial thickness of the shell. As relation (1) is an empirical one it is impossible to decide whether the value of ρ which enters into equation (4) is the mean density of the shell or not. In order to make the analysis quite general let it be the density of the shell weighted by a certain factor $g(a)$ which depends on some property of the whole configuration, for example if the magnetic effect of a shell depend upon the gravitational field at that shell we find

$$dP = \frac{4\pi}{3} \beta \frac{G^{\frac{1}{2}}}{c} \rho(a)g(a)a^4 \omega da. \quad \dots\dots(5)$$

Thus relation (1) now becomes

$$P = \frac{4\pi}{15} \beta \frac{G^{\frac{1}{2}}}{c} R^5 \omega D(R), \quad \dots\dots(6)$$

where $D(R) = \int_0^R \rho(a)g(a)a^4 da / \int_0^R a^4 da$, a mean density weighted by the factor $g(a)$. By placing $g(a) = 1$ we obtain the relation

$$P = \frac{4\pi}{15} \beta \frac{G^{\frac{1}{2}}}{c} k R^5 \omega \rho_m, \quad \dots\dots(7)$$

where $k = I/I_0$ where I is the moment of inertia of a given centrally condensed sphere and I_0 that of a sphere of the same radius and angular velocity and of a

uniform density equal to the mean density of the centrally condensed sphere. The quantity on the right hand of equation (6) is a constant $\beta G^{\frac{1}{2}}/2c$ multiplied by the angular momentum of the sphere, and Blackett obtains values for β as given in table 1 for different astronomical bodies using experimental values of P and U and an assumed value for k . The uncertainties in the values of R , ω and P for the Sun and 78 Virginis do not allow us to decide exactly how density should enter into the formula (1), i.e. to determine the function $g(a)$, as the range of densities between the three bodies is rather small.

It is possible that the magnetic effect of the shell at radius a may depend upon the gravitational field at that shell, in which case the ρ in formula (3) will have to be replaced by a quantity dependent partly on the density of the shell itself and partly on the average density of the matter within the shell.

Table 1

	Average ρ	Values of β	
		Assuming uniform density	Inserting constant k
Earth	5.53	0.26	0.30
Sun	1.39	0.18	1.14
78 Virginis	0.4	0.19	1.16

Though we cannot assume that on a distributed theory the magnetic field is irrotational within the matter of the rotating body, we can assume that outside the shell of matter the magnetic field may be represented by a scalar potential. In order to calculate the field components at depth d we imagine the Earth divided into a thick shell bounded by the surfaces $r=R$ and $r=R-d$ and a sphere of radius $R-d$, the two being separated by an infinitesimal spherical shell free of matter. The part of the field components at depth d due to the sphere may be written down from (6) and the part due to the thick shell may be obtained by differentiating within the spherical shell free of matter the scalar potential at depth d due to the thick shell which is obtained by integrating the contributions from such elementary shells.

The potential at (r, θ) for $r < R-d$ due to a shell extending from $r=R-d$ to $r=R$ will be

$$-\frac{8\pi}{3} \beta \frac{G^{\frac{1}{2}}}{c} \omega r \cos \theta \int_{R-d}^R \rho(a) g(a) a da,$$

where the integral will depend on the density of the surface rocks.

The field intensities at depth d may now be derived.

The radial component $H_d^r = [\partial \phi_d / \partial r]_{r=R-d} - 2[P(R-d)/(R-d)^3] \cos \theta$, where $P(R-d)$ is the magnetic moment of the sphere radius $R-d$ is

$$H_d^r = H_0^r \left[\frac{D(R-d)}{D(R)} \left(1 - \frac{d}{R} \right)^2 + \frac{5}{2} \frac{\Delta(d)}{D(R)} \left(2 \frac{d}{R} - \frac{d^2}{R^2} \right) \right], \quad \dots \dots (8)$$

where $\Delta(d) = \int_{R-d}^R \rho(a) g(a) a da / \int_{R-d}^R a da$, a weighted surface density.

Similarly the horizontal component of the magnetic field at depth d is

$$\begin{aligned} H_d^\theta &= [\partial \phi_d / r \partial \theta]_{r=R-d} - P(R-d) \sin \theta / (R-d)^3 \\ &= H_0^\theta \left[\frac{D(R-d)}{D(R)} \left(1 - \frac{d}{R} \right)^2 - 5 \frac{\Delta(d)}{D(R)} \left(2 \frac{d}{R} - \frac{d^2}{R^2} \right) \right]. \quad \dots \dots (9) \end{aligned}$$

For a density distribution not differing greatly from a uniform one

$$D(R-d) = D(R) = \rho_m \quad \text{and} \quad \Delta(d) = \rho_s,$$

where ρ_m and ρ_s are respectively the mean density of the Earth and the density of the surface rocks.

Equation (9) then becomes

$$H_d^0 = H_0^0 \{1 - (2 + 10 \rho_s / \rho_m)(d/R)\}, \quad \dots\dots (10)$$

which is the formula obtained by the author, communicated to and quoted by Hales and Gough (1947). It fits the case of the Earth reasonably well down to depths of 1 mile under land. As might be expected for a theory which assumes that the whole of the Earth is contributing to the magnetism, H^0 should decrease with depth, while on the core theories it should increase with depth. Thus there emerges the possibility of being able to determine by variation with depth experiments which theory of the Earth's magnetism is true.

In Great Britain the various differences would be

$$\left. \begin{aligned} H_d^r - H_0^r &= +25 \gamma \text{ per 4000 ft.} \\ H_d^0 - H_0^0 &= +11 \gamma \text{ per 4000 ft.} \end{aligned} \right\} \text{core theory,}$$

$$\left. \begin{aligned} H_d^r - H_0^r &= +25 \gamma \text{ per 4000 ft.} \\ H_d^0 - H_0^0 &= -23 \gamma \text{ per 4000 ft.} \end{aligned} \right\} \text{distributed theory.}$$

Some preliminary measurements have been made by Hales and Gough (1947) on the suggestion of Dr. E. C. Bullard in South Africa and by Benson, Goldsack and Runcorn (see Chapman 1948) in Lancashire. These appear to support the distributed theory.

Let us now examine the exact formula of the field variation for $g(a) = 1$ for small depths.

In this case $D(R) = \int_0^R \rho(a) a^4 da / \int_0^R a^4 da = k\rho_m$, and for small d/R ,

$$\begin{aligned} \frac{1}{5} R^5 D(R) &= \int_0^{R-d} \rho(a) a^4 da + \frac{1}{5} \rho_s \{R^5 - (R-d)^5\} \\ R^5 D(R) &= (R-d)^5 D(R-d) + \rho_s \{R^5 - (R-d)^5\}. \end{aligned}$$

Then

$$\begin{aligned} H_d^r &= H_0^r \left[\left(1 + \frac{3d}{R}\right) \left(1 - \frac{5d}{R} \frac{\rho_s}{k\rho_m}\right) + \frac{5}{k} \frac{\rho_s}{\rho_m} \frac{d}{R} \right] \\ &= H_0^r [1 + 3d/R] \text{ for small } d/R \end{aligned} \quad \dots\dots (11)$$

and

$$\begin{aligned} H_d^0 &= H_0^0 \left[\left(1 + \frac{3d}{R}\right) \left(1 - \frac{5d\rho_s}{Rk\rho_m}\right) - \frac{10\rho_s}{k\rho_m} \frac{R}{d} \right] \\ &= H_0^0 [1 + 3d/R - 15\rho_s/k\rho_m(d/R)] \text{ for small } d/R, \quad \dots\dots (12) \end{aligned}$$

which is the formula given by Chapman (Chapman and Runcorn 1948). Thus H^0 decreases under land on the Earth according to the formula

$$H_d^0 = H_0^0 [1 - 6.5d/R] \text{ assuming } \rho_s = 3,$$

and under sea according to the formula

$$H_d^0 = H_0^0 [1 - 0.2d/R].$$

If $g(a)$ is appreciably different from 1 this should be detectable experimentally and might be of importance in explaining relation (1).

The variation in H^θ and H^r within the Earth may be computed from formulae (8) and (9) using the density distribution inside the Earth known from seismological data. In figure 1 the density distribution is taken from the mean of the two

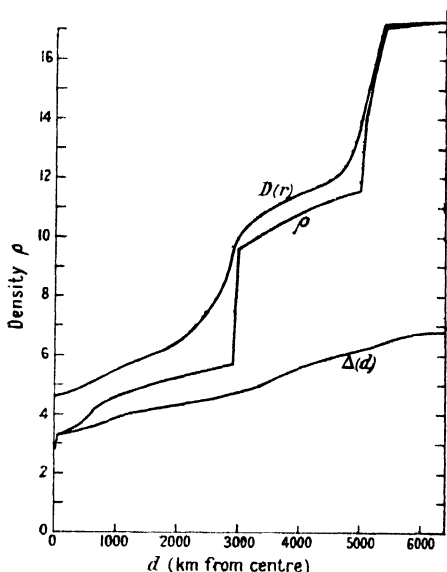
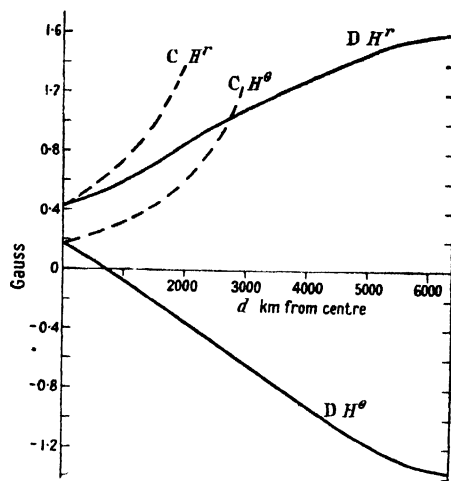


Figure 1. Variation of densities within the Earth.



C = Core theory. D = Distributed theory.

Figure 2. Variation of H^r and H^θ within the Earth for $H_0^\theta = 0.18$ gauss; $H_0^r = 0.42$ gauss.

sets given by Bullen (1947, p. 218). In formulae (8) and (9), $g(a)$ is taken equal to 1 for simplicity. The results are plotted in figure 2. It will be seen that H^θ reverses at a depth of about 700 miles below the surface.

§ 7. A PROPOSED FURTHER INVESTIGATION OF MASS AS VIRTUAL CURRENT FLOW

If the measurements of the Earth's field at depth finally give a positive result in favour of a distributed theory such as Blackett's, it is important to make further investigations as to the magnetic effect of the matter in the surface shell S.

The boundary condition selected to derive the variation of depth formula from relation (1) implies that the shell must resemble a flow of electric current, and it is important to decide if this is so. Consider an elementary spherical shell: we have shown in § 5 that there is a discontinuity of potential at the shell—this implies that the latter is equivalent to a magnetic shell of moment M per unit area where

$$4\pi M = (4G^1/3c)\pi\beta\omega\rho a^2 da \cos\theta + (16G^1/3c)\pi\beta\omega\rho a^2 da \cos\theta,$$

$$M = \beta(G^1/c)\omega\rho a^2 \cos\theta da.$$

Suppose each ring of matter in the shell is equivalent to a magnetic shell of strength m per unit cross-section of ring (Chapman and Bartels 1940, chap. 17). Then the shell may be taken over the cap of the shell on the side remote from the north pole and bounded by the ring at co-latitude θ .

Hence the moment per unit area $M = \int_0^\theta mad \, d\theta$, and

$$m = -\beta G^{\frac{1}{2}} \omega \rho a \sin \theta / c = -\beta G^{\frac{1}{2}} \rho v / c \quad \dots\dots (13)$$

is the peripheral velocity of the ring.

This is equivalent to H. A. Wilson's hypothesis that a mass element m gm. when moving with a velocity v gives at a distance r a magnetic field

$$H = -\beta(G^{\frac{1}{2}}m/cr^3)\mathbf{v} \wedge \mathbf{r}$$

which he postulated to explain a relation similar to (1).

Wilson (1923) disproved this hypothesis by experiment for linear motion and Blackett (1947) showed that it was untenable on theoretical grounds. However, Blackett suggests that by replacing v by $\omega \wedge R$ it may be possible to retain Wilson's hypothesis for rotational motion but not for translational motion (cf. equation (13)).

In any case if the boundary condition selected in §3 which leads to the derivation of equations (8) and (9) is correct we should expect a large and sufficiently long mountain range or deep gorge running E-W to produce a magnetic field of a similar type to that of a linear current. A calculation based on equation (12) using the value of β for the Earth given in table 1 shows that a gorge 4000 ft. (1.3 km.) deep and 3 km. wide will produce a peak field of about 13γ at its edge (figure 3).

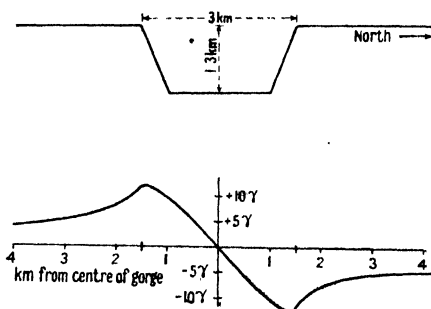


Figure 3. Magnetic effect of gorges on vertical component of full intensity.

In the Northern Hemisphere the vertical component of the Earth's field would be increased by 13γ on the south side of the gorge and decreased by 13γ on the north side. The effect will fall off inversely as the distance from the centre of the gorge. To make such an experiment with success the rock round the gorge would have to be of low susceptibility (about 10^{-5} gauss) and the strata should lie horizontal. This would exclude most natural features except those gorges formed by erosion in sedimentary rock. It might well be that airborne magnetometers would be very useful in investigations of this effect.

ACKNOWLEDGMENTS

The author is indebted to Professor L. Rosenfeld for initial checking of the calculations, to Professors E. C. Bullard and S. Chapman for valuable discussion, to the latter for contributing an Appendix to this paper and to Professor P. M. S. Blackett for much encouragement and help.

REFERENCES

- BLACKETT, P. M. S., 1947, *Nature, Lond.*, **159**, 658.
 BULLARD, E. C., unpublished.
 BULLEN, K. E., 1947, *Introduction to the Theory of Seismology* (Cambridge: University Press).
 CHAPMAN, S., 1948, *Nature, Lond.*, **161**, 462.
 CHAPMAN, S., and BARTELS, J., 1940, *Geomagnetism* (Oxford: Clarendon Press).

- CHAPMAN, S., and RUNCORN, S. K., 1948, *Nature, Lond.*, **161**, 52.
 ELSASSER, W. M., 1939, *Phys. Rev.*, **55**, 489 ; 1941, *Ibid.*, **60**, 876.
 FLEMING, J. A. (Ed.), 1940, *Terrestrial Magnetism and Electricity* (London and New York : McGraw Hill).
 FRENKEL, J., 1945, *C.R. Acad. Sci. U.S.S.R.*, **49**, 98.
 HAALCK, H., 1937, *Z. Phys.*, **105**, 81 ; 1938, *Gerlands Beitr. Geophys.*, **52**, 243.
 HALES, A. L., and GOUGH, D. I., 1947, *Nature, Lond.*, **160**, 746.
 SCHLOMKA, T., 1933, *Gerlands Beitr. Geophys.*, **38**, 357.
 SCHUSTER, A., 1912, *Proc. Phys. Soc.*, **24**, 121.
 STRATTON, J. A., 1941, *Electromagnetic Theory* (London and New York : McGraw Hill).
 SUTHERLAND, W., 1900, *Terr. Magn. Atmos. Elect.*, **5**, 73 ; 1903, *Ibid.*, **8**, 49.
 WILSON, H. A., 1923, *Proc. Roy. Soc. A*, **104**, 451.

APPENDIX

By SYDNEY CHAPMAN

Mr. Runcorn has kindly invited me to append my deduction of the formula quoted in his §6 and by Chapman and Runcorn (1948, equation (3)); this is briefly as follows. I hope to communicate to *Annales de Géophysique* a fuller treatment of this and other questions related to the Blackett hypothesis of the earth's magnetic field.

(a) From general electromagnetic theory we know that the field of magnetic induction (this and the field of magnetic intensity will here be denoted by b and h , and $b = h + 4\pi I$) for a body magnetized with uniform intensity I is the same as the field of magnetic intensity H of a distribution of current of surface intensity $I \wedge n$ over the boundary of the body (here n denotes the outward normal unit vector at the surface point considered, and $\wedge$ denotes the vector product sign).

For such a body of spherical form, with radius R , the field b at inside points is $(8/3)\pi I$, and at outside points it is $-\text{grad } V$, where $V = M \cos \theta / r^2$ at a point whose polar coordinates are r, θ , where $M = (4/3)\pi R^3 I$. This is therefore the field H of a surface current of intensity $I \sin \theta$ flowing round the parallels of magnetic latitude.

(b) Blackett in effect supposes that, in a solid of revolution a rotatory flow of mass has the magnetic effect of a proportional electric current flow. This current flow, in a sphere of radius a rotating with angular speed ω , and with density ρ dependent only on R , is $K\rho\omega R \sin \theta$ at radius R , where K is a constant written by Blackett in the form $\beta G^4/2c$.

Such a volume current distribution can be supposed subdivided into thin shells (of radii $R, R + dR$). Such a shell can be treated as a surface current flow of density $K\rho\omega R \sin \theta dR$ over the sphere of radius R ; its magnetic field dh is the same as the b field of this sphere uniformly magnetized with intensity $I = K\rho\omega R dR$, parallel to the axis of rotation. To obtain the whole field H it is necessary to integrate the b fields for all such contributions. At a point at radius r within the sphere, H is the integral of b according to the formula $(8/3)\pi I$ for all the shells of radius $0 \leq R \leq r$, and according to the second formula in (a), for $r \leq R \leq a$.

It is convenient to express the results by the vertical (radial) and horizontal components H^r, H^θ , as follows:

$$H^r = (8/3)\pi K\omega \cos \theta \left\{ \int_0^r \rho R^4 dR / r^3 + \int_r^a \rho R dR \right\},$$

$$H^\theta = (4/3)\pi K\omega \sin \theta \left\{ \int_0^r \rho R^4 dR / r^3 - 2 \int_r^a \rho R dR \right\}.$$

(c) At the surface $r=a$ (here indicated by the suffix 0) these components have the values indicated by

$$H_0^r = 2H_0^\theta \cot \theta,$$

$$H_0^\theta = (4/3)\pi K\omega \sin \theta \int_0^a \rho R^4 dR/a^3 = \frac{1}{2}K\omega(I/a^3) \sin \theta,$$

where I denotes the moment of inertia of the sphere. Taking this moment to be kI_0 , where I_0 denotes the moment $(8\pi ka^5\rho_m/15)$ for a uniform sphere of radius a and the same mean density ρ_m ,

$$H_0^\theta = (4/15)\pi kK\omega a^2\rho_m \sin \theta.$$

(d) The rates of radial variation of H^r and H^θ are obtained by differentiating the expressions given in (b) with respect to r . The results are specially simple at the surface $r=a$ (again indicated by suffix 0), as follows:

$$\begin{aligned} \left(\frac{dH^r}{dr}\right)_0 &= -\frac{8\pi K\omega \cos \theta}{a^4} \int_0^a \rho R^4 dR = -3 \frac{H_0^r}{a} \\ \left(\frac{dH^\theta}{dr}\right)_0 &= \frac{4}{3} \pi K\omega \sin \theta \left\{ 3\rho_s a - \frac{3}{a^4} \int_0^a \rho R^4 dR \right\} \\ &= -3 \frac{H_0^\theta}{a} \left(\frac{5\rho_s}{k\rho_m} - 1 \right), \end{aligned}$$

where ρ_s denotes the surface density.

The first of these two equations indicates that at the surface the rate of variation of the vertical component with respect to radius (or depth) is the same as for a core theory; the second equation is equivalent to my equation (3) in the article by Chapman and Runcorn (1947).

A Numerical Fourier-analysis Method for the Correction of Widths and Shapes of Lines on X-ray Powder Photographs

By A. R. STOKES

Wheatstone Physics Laboratory, King's College, University of London*

Communicated by J. T. Randall; MS. received 15 April 1948

ABSTRACT. A method is derived, using Fourier analysis, for finding the corrected distribution of intensity across an x-ray diffraction line, and the procedure is illustrated by a numerical example. The method may also be of use in spectrum analysis and statistical problems.

§ 1. INTRODUCTION

A STUDY of the widths of the lines on x-ray (Debye-Scherrer) powder photographs can yield valuable information as to the particle size, state of strain, or "mistake structure" in a specimen (see for instance Jones (1938), Edwards, Lipson and Wilson (1942), Megaw and Stokes (1945), and Birks and Friedman (1946)). But the line widths obtained by photometry

* An account of work done mainly at the Cavendish Laboratory, Cambridge.

are not of much value until allowance has been made for the additional broadening inherent in the method of photography, due mainly to the specimen size and to the slight spread of wavelengths (including the $K\alpha$ doublet) in the incident radiation. This "instrumental" broadening is very seldom negligible.

Kochendörfer (1937) and Jones (1938) have described corrections for instrumental broadening; the former calculated the broadening to be expected from the apparatus while the latter observed the broadening actually obtained from the apparatus when a specimen of powdered metal, having strain-free crystals large enough to contribute no broadening to the lines, was used.

Both these methods, however, involve some assumption as to the intensity distribution across the lines which the specimen would give if instrumental broadening were absent. Kochendörfer used a theoretical line shape derived by Laue, which is not necessarily appropriate in all cases; in Jones' method a certain amount of trial and error is needed before the right line profile can be obtained. It seems desirable, therefore, to devise a direct method whereby the line profile which would be obtained in the absence of any instrumental broadening can be deduced from the experimental line profiles. Further, the methods referred to above give only the corrected line *breadths*, whereas useful information may often be obtained from line *shapes*. The breadth of a line (whether it be the half-peak width, the "integral breadth", or some other measure) is in fact only *one* parameter (though often a very valuable one) defining the distribution of x-ray intensity across the line.

§ 2. MATHEMATICAL BASIS OF THE METHOD

The observed distribution of x-ray intensity across a line is related to the distribution due to instrumental broadening and that due to small particle size etc., in the specimen, by the equation

$$h(x) = \int_{-\infty}^{\infty} f(y)g(x-y)dy \quad \dots\dots(1)$$

where $f(x)$ is the intensity which would be recorded, at a point distance x measured along the film from some fixed point, in the absence of instrumental broadening, $g(x)$ is similarly the intensity recorded at x from a specimen giving negligible broadening in itself, so that all broadening is instrumental, and $h(x)$ is the intensity actually observed, when the broadening is due to both causes. This relation is proved by Jones (1938). The functions $h(x)$ and $g(x)$ are known (being given in numerical form from intensity measurements) while $f(x)$ is an unknown function, which Jones found by trial and error, but which we want to deduce from g and h .

In practice we generally measure g and h at discrete intervals (say at every 0.1 mm. of film) instead of as continuous functions, and it is therefore convenient to replace the integral in (1) by a summation

$$h(x) = \Sigma f(y)g(x-y)\delta y \quad \dots\dots(2)$$

where $\delta y = 0.1$ mm.: this is very nearly equivalent to the integral form (1). We see that the relation in this form constitutes a set of linear simultaneous equations, one equation for each value of x for which $h(x)$ is measured, with the values of $f(x)$ as the unknowns. As there are in general about 20 to 50 of these equations, it would be too laborious to solve them by the ordinary method of

elimination. An iteration method (e.g. Whittaker and Robinson 1944 a) might be applied, but as Shull (1946) has pointed out, one can use Fourier analysis to obtain the solution. We take a range of values of x from $-a/2$ to $+a/2$ outside which $g(x)$ and $h(x)$ are zero (that is, outside which the intensity can be considered to have fallen to its background value), and we obtain Fourier series for $g(x)$, $h(x)$ and $f(x)$ in this range :

$$\left. \begin{aligned} f(x) &= \sum_{-\infty}^{\infty} F(t) \exp(-2\pi ixt/a) \\ g(x) &= \sum G(t) \exp(-2\pi ixt/a) \\ h(x) &= \sum H(t) \exp(-2\pi ixt/a) \end{aligned} \right\} \dots\dots(3)$$

where $t = 0, \pm 1, \pm 2$ etc., and F, G , and H , are complex coefficients whose values are given by equations

$$F(t) = \frac{1}{a} \int_{-a/2}^{+a/2} f(x) \cdot \exp(2\pi ixt/a) dx \text{ etc.} \dots\dots(4)$$

We now substitute the expressions for f and g given by equation (3) into equation (1), and alter the limits of integration to $-a/2$ and $+a/2$ (for f and g are zero outside these limits) and we obtain

$$\begin{aligned} h(x) &= \int_{-a/2}^{+a/2} f(y) g(x-y) dy \\ &= \int \sum_t \sum_{t'} F(t) G(t') \exp(-2\pi iyt/a) \exp\{-2\pi i(x-y)t'/a\} dy \\ &= \sum_t \sum_{t'} F(t) G(t') \int_{-a/2}^{+a/2} \exp\{-2\pi iy(t-t')/a\} \exp(-2\pi ixt'/a) dy. \end{aligned}$$

The integral of $\exp\{-2\pi iy(t-t')/a\}$ is zero if $t \neq t'$ and equal to a if $t = t'$; thus the only terms left in the summation are those with $t = t'$, and the series reduces to

$$h(x) = a \sum_t F(t) G(t) \exp\{-2\pi ixt/a\}.$$

Comparing this with the third equation of (3), we obtain

$$H(t) = a F(t) G(t)$$

or

$$F(t) = H(t)/aG(t). \dots\dots(5)$$

By substituting this in the first of equations (3) we obtain

$$f(x) = \sum_t \frac{H(t)}{aG(t)} \exp\{(-2\pi ixt)/a\}. \dots\dots(6)$$

In practice we omit the factor $1/a$ since it is a constant and does not affect the shape of the resulting curve, but only its height. We can therefore express the procedure for finding $f(x)$ as follows. Find the Fourier components of the distributions of intensity $h(x)$ and $g(x)$ of the observed broad line and sharp line respectively : divide each Fourier component of $h(x)$ by the corresponding component of $g(x)$, and use the resulting quotients in a Fourier synthesis to find $f(x)$. It should perhaps be pointed out that this procedure eliminates broadening due to $K\alpha$ doubling as well as other forms of broadening; there is no need to deal separately with $K\alpha$ doubling as in Jones' method. The Fourier components are, of course,

complex numbers, and the method of dealing numerically with these complex quantities will be explained subsequently. The expression (1) for h in terms of f and g is known to mathematicians as the "fold" of f and g ; it seems therefore appropriate to call f the "unfold" of h by g , and to call the process by which f is obtained, "unfolding".

Shull (1946) has used equation (6) to show that, if $h(x)$ and $g(x)$ are fitted to certain functions made up of Gaussian error functions, then the unfold can be found in terms of similar functions, and its breadth can be found in terms of certain parameters of the original functions.

This procedure evidently gives us a fairly rapid method of obtaining reasonably reliable values of the corrected breadths, but it seems somewhat undesirable to fit the observed intensity curves to Gaussian error functions, partly because in so doing one introduces an element of arbitrary rigidity into the treatment which may lead to a deceptive appearance of certainty about the results, and partly because theory does not give functions of this form for the intensity distributions, at any rate from small crystals (Stokes and Wilson 1942, 1944). Also it appears that difficulties would be encountered when $K\alpha$ doubling causes appreciable asymmetry in the line profiles, since the functions that Shull uses are all symmetrical. In view of the existence of Lipson-Beevers strips (Lipson and Beevers 1936) and other devices for enabling Fourier analyses to be done rapidly (e.g. Hägg and Laurent 1946) it is not a very formidable task to evaluate the functions $f(x)$ numerically from equations (4) and (6).

§ 3. EXPERIMENTAL MATERIAL AND METHOD OF COMPUTATION

The method was applied to investigate the shapes and breadths of reflections from cold-worked copper. A specimen of copper filings in a sealed thin-walled silica tube was photographed in a high-temperature x-ray powder camera at room temperature; the temperature was then raised to about 400° C. for an hour to anneal the copper filings, and the specimen was again photographed at room temperature, giving a set of lines whose breadths were due almost entirely to instrumental broadening. The photographs were microphotometered, and the functions $h(x)$ and $g(x)$ were obtained, for each line, from the microphotometer readings for the respective photographs, by subtracting the background intensity from the observed readings of intensity. The microphotometer gave readings of density at intervals of 0.1 mm. Corrections were made where necessary for any departure from a linear density-intensity relation, as shown by a step-wedge calibration.

The Fourier components $H(t)$ and $G(t)$ of the functions $h(x)$ and $g(x)$ were obtained by the use of Lipson-Beevers strips. On these are tabulated values of $A \cos nx$ and $B \sin nx$ for A and B equal to integers from -99 to $+99$, $n=0, 1, 2, \dots, 20$, and $x=0^\circ, 6^\circ, 12^\circ, \dots, 90^\circ$, or $0, 1, 2, \dots, 15 \times 2\pi/60$ radians. Thus they facilitate the evaluation of sums like $\sum_x g(x) \cos 2\pi xt/60$ and $\sum_x g(x) \sin 2\pi xt/60$, where $g(x)$ is given at $x=0, \pm 1, \pm 2, \dots$ (in arbitrary units of length) and $t=0, 1, 2, \dots$

If $G_r(t)$ and $G_i(t)$ are the real and imaginary parts of $G(t)$, then from equation (4)

$$G_r(t) = \frac{1}{a} \int_{-a/2}^{a/2} g(x) \cos(2\pi xt/a) dx \quad \text{and} \quad G_i(t) = \frac{1}{a} \int_{-a/2}^{a/2} g(x) \sin(2\pi xt/a) dx.$$

To evaluate these integrals numerically we divide the range $-a/2$ to $a/2$ into 60 intervals, and sum the integrands at these intervals. For convenience we let our unit of length be such that $a = 60$ units, so that x is taken at unit intervals. Then

$$G_r(t) = \frac{1}{60} \sum_{-30}^{30} g(x) \cos 2\pi xt/60, \quad \text{and} \quad G_i(t) = \frac{1}{60} \sum_{-30}^{30} g(x) \sin 2\pi xt/60, \quad \dots\dots(7)$$

which may be evaluated by means of Lipson-Beevers strips as described above. The validity of this change from an integral to a sum depends upon whether the intervals chosen are small enough compared with the scale of the detail in the $g(x)$ curve; we shall return to this point later.

Having obtained $G_r(t)$, $G_i(t)$, and similarly $H_r(t)$, $H_i(t)$, we now have

$$F(t) = \frac{H(t)}{G(t)} = \frac{H_r(t) + iH_i(t)}{G_r(t) + iG_i(t)} = \frac{(H_r + iH_i) \cdot (G_r - iG_i)}{G_r^2 + G_i^2};$$

hence
$$F_r = \frac{H_r G_r + H_i G_i}{G_r^2 + G_i^2}; \quad F_i = \frac{H_i G_r - H_r G_i}{G_r^2 + G_i^2},$$

from which we can calculate $F_r(t)$ and $F_i(t)$. Alternatively, a complex-number slide-rule * (Du Mond 1924) might be used.

Finally, we have

$$\begin{aligned} f(x) &= \sum_t F(t) \exp [2\pi ixt/a] \\ &= \sum_t \{F_r(t) + iF_i(t)\} (\cos 2\pi xt/60 - i \sin 2\pi xt/60) \\ &= \sum_t F_r(t) \cos 2\pi xt/60 + \sum_t F_i(t) \sin 2\pi xt/60, \end{aligned}$$

the other terms vanishing since $F(-t)$ is the conjugate complex of $F(t)$, and therefore $F_r(-t) = F_r(t)$, and $F_i(-t) = -F_i(t)$.

The procedure is illustrated by the process of unfolding one of the lines from the photographs of cold-worked and annealed copper, given in tables 1 to 4.

Table 1. $h(x)$

x	$h(x)$	x	$h(x)$	x	$h(x)$	x	$h(x)$
-20	3	-10	33	0	277	10	29
-19	4	-9	42	1	243	11	26
-18	5	-8	50	2	233	12	21
-17	7	-7	60	3	208	13	18
-16	9	-6	82	4	177	14	16
-15	11	-5	102	5	147	15	14
-14	13	-4	140	6	125	16	12
-13	18	-3	194	7	83	17	10
-12	22	-2	237	8	70	18	8
-11	27	-1	267	9	39	19	7
						20	7

(x in tenth-millimetres.)

Each value of $h(x)$ as given here was subsequently multiplied by 0.2 for convenience in using Lipson-Beevers strips.

* The general principle of this slide-rule is that, just as real numbers can be multiplied by the addition of their logarithms plotted on a one-dimensional scale, so complex numbers may be multiplied by the vector addition of their logarithms plotted on a two-dimensional sheet. The logarithms are evaluated from the formula

$$\ln(x + iy) = \frac{1}{2} \ln(x^2 + y^2) + i \tan^{-1}(y/x).$$

The real part of this is plotted as abscissa and the imaginary part as ordinate; division can then be effected by subtracting the vectors found by joining points on this plot to the origin.

Table 2. $g(x)$

x	$g(x)$	x	$g(x)$	x	$g(x)$	x	$g(x)$
-13	0	-7	9	0	475	7	60
-12	0	-6	12	1	371	8	19
-11	0	-5	17	2	174	9	12
-10	2	-4	21	3	151	10	8
-9	4	-3	51	4	227	11	5
-8	6	-2	161	5	260	12	3
		-1	369	6	160	13	1

(x in tenth-millimetres.)

Each value of $g(x)$ as given here was subsequently multiplied by 0.1 for convenience in using Lipson-Beevers strips.

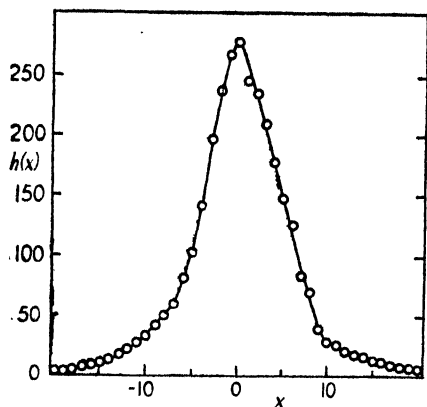
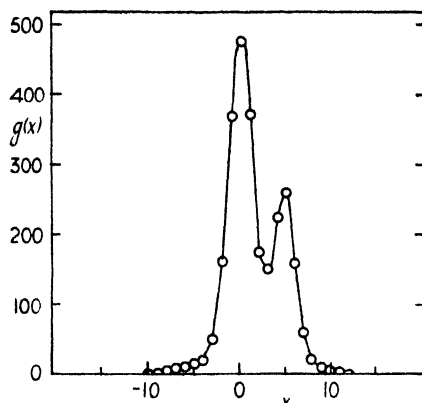
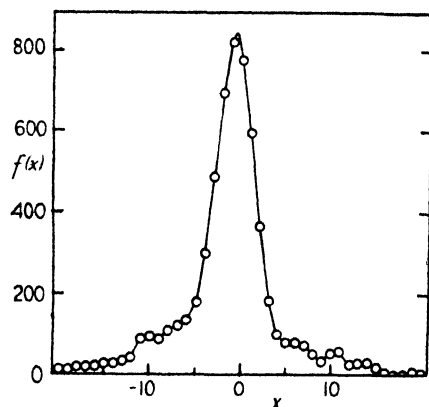
Table 3. Fourier components of $h(x)$, $g(x)$, and $f(x)$

t	$H_r(t)$	$H_l(t)$	$G_r(t)$	$G_l(t)$	$F_r(t)$	$F_l(t)$
0	619	0	259	0	2.40	0
1	517	16	243	37	2.10	-0.25
2	334	30	203	63	1.55	-0.35
3	205	30	152	70	1.20	-0.35
4	125	22	102	61	1.00	-0.38
5	63	7	69	36	0.75	-0.30
6	33	-10	53	7	0.60	-0.30
7	18	-22	50	-18	0.48	-0.28
8	10	-23	55	-33	0.35	-0.25
9	12	-17	61	-37	0.28	-0.18
10	6	-11	67	-32	0.15	-0.10
11	2	-5	66	-21	0.05	-0.05
12	2	-4	62	-9	0.05	-0.05
13	-4	-5	51	1	—	—
14	-2	-7	39	8	—	—
15			24	10		
16			14	7		
17			7	3		
18			5	2		
19			2	0		
20			4	0		

Each value of $F_r(t)$ and $F_l(t)$ was subsequently multiplied by 40 for convenience in using Lipson-Beevers strips.

Table 4. $f(x)$, from synthesis of $F_r(t)$ and $F_l(t)$

x	$f(x)$	x	$f(x)$	x	$f(x)$	x	$f(x)$
-20	15	-10	93	0	780	10	53
-19	13	-9	85	1	599	11	59
-18	18	-8	107	2	369	12	26
-17	19	-7	121	3	185	13	29
-16	23	-6	135	4	100	14	29
-15	28	-5	178	5	80	15	20
-14	29	-4	298	6	79	16	7
-13	35	-3	489	7	73	17	1
-12	40	-2	697	8	53	18	4
-11	87	-1	821	9	33	19	9
						20	7

Figure 1. $h(x)$.Figure 2. $g(x)$.Figure 3. $f(x)$.

The shapes of the curves $h(x)$, $g(x)$, and $f(x)$ are shown in figures 1, 2 and 3.

We can investigate the validity of replacing the integrals in (4) by summations in (7) as follows. We have $g(x) = \sum_t G(t) \exp(-2\pi ixt/a)$, by (3), and we wish to find how well $G'(t) \equiv a^{-1} \sum_{-a/2}^{a/2} g(x) \exp(2\pi ixt/a)$ approximates to

$$G(t) \equiv a^{-1} \int_{-a/2}^{a/2} g(x) \exp(2\pi ixt/a) dx,$$

x being taken at unit intervals in the summation. By (3) we have

$$\begin{aligned} G'(t) &= a^{-1} \sum_{x'} \sum_{t'} G(t') \exp(-2\pi ixt'/a) \exp(2\pi ixt/a) \\ &= a^{-1} \sum_{t'} G(t') \left[\sum_x \exp(2\pi ix(t-t')/a) \right]. \end{aligned}$$

The sum within the square brackets is the sum of a geometric series, which is zero if $(t-t')/a$ is not an integer, and equal to a if $(t-t')/a$ is an integer. Thus $G'(t)$ is equal to the sum of all the values of $G(t')$ for $t-t'$ equal to an integral multiple of a , that is,

$$\begin{aligned} G'(t) &= G(t) + G(t+a) + G(t+2a) + \dots \\ &\quad + G(t-a) + G(t-2a) + \dots \end{aligned} \quad \dots\dots(8)$$

We see from the numerical example that H and G fall to zero (apart from small fluctuations due to experimental error) as t increases. Now if $G(t)$ is zero for all values of t outside the range $-a/2$ to $+a/2$, then not more than one term of the series in equation (8) can be non-zero at once, and therefore $G'(t) = G(t)$ in the range $-a/2 < t < a/2$.

The condition that the summations (7) should be good representations of the Fourier components is therefore that $G(t)$ should fall sensibly to zero for values of t outside a range not larger than the range $-a/2$ to $+a/2$. In the numerical example above, $H(t)$ and $G(t)$ become nearly zero at $t=12$ and $t=18$ respectively. Since $a=60$, these values of t fall within the required range, -30 to $+30$, and therefore the values obtained by summation are the same as the true values which would be obtained by integration over the whole range of t in which we are interested. If this were not so, we should have to use a greater number of divisions. From the above reasoning, the number of divisions must always exceed twice the value of t for the Fourier component of lowest frequency which is sensibly zero.

§ 4. EFFECT OF EXPERIMENTAL ERRORS

In any set of microphotometry readings the measurements of density are subject to errors. We shall investigate here how the errors in the unfolded curves depend on those of the observed curves.

We assume (a) that the probable error of any one photometer reading is the same as that of any other, and (b) that the errors of any two readings, even adjacent ones, are independent. This may be taken as true when the grain-size of the film (which is the main cause of error in microphotometry) is small compared with the 0.1 mm. distance between successive readings.

The probable errors in the Fourier components $G(t)$ and $H(t)$ can be obtained as follows. If $\Delta g(x)$ is the error in $g(x)$, then the error in $G(t)$ is

$$\Delta G(t) = a^{-1} \sum_x \Delta g(x) \exp(2\pi i x t)$$

and the probable error of $G(t)$ can be assessed from the mean square value of the modulus of the error :

$$\begin{aligned} |\overline{\Delta G(t)}|^2 &= \overline{\Delta G(t) \Delta G(t)^*} \\ &= a^{-2} \sum_x \Delta g(x) \exp(2\pi i x t/a) \sum_x \Delta g(x) \exp(-2\pi i x t/a) \\ &= a^{-2} \sum_{x_1, x_2} \overline{\Delta g(x_1) \Delta g(x_2)} \exp(2\pi i (x_1 - x_2)t/a), \end{aligned}$$

the "bar" above an expression denoting that the mean value of that expression is to be taken. But by our assumptions (a) and (b) above $\overline{\Delta g(x_1) \Delta g(x_2)} = 0$ for x_1 and x_2 different, and $\overline{\Delta g(x)^2} = \text{const.} = \Delta^2$, say. Hence

$$|\overline{\Delta G(t)}|^2 = a^{-2} \sum_x \Delta g(x)^2 = a^{-2} \cdot a \Delta^2 = \Delta^2/a$$

(Whittaker and Robinson 1944 b), since a is the number of readings over which the summation is taken. The mean square error in $F(t)$ is given by

$$\begin{aligned} |\overline{\Delta F(t)}|^2 &= |\overline{\Delta(H/G)}|^2 = \overline{(G\Delta H - H\Delta G)^2/G^2} \\ &= (G^2 \overline{\Delta H^2} + H^2 \overline{\Delta G^2})/G^4 = \{(G^2 + H^2)/G^4\} \cdot (\Delta^2/a). \end{aligned}$$

Since G^4 occurs in the denominator of this expression, the probable error of $F(t)$ increases rapidly as $G(t)$ becomes smaller. Reference to the values of $G(t)$ given in table 2 will show that $|G(t)|$ decreases as t increases; hence the error in $F(t)$ will increase as t increases. This increase is not serious, however, provided $H(t)$ has fallen sensibly to zero before $G(t)$ has decreased very much from its maximum value, so that terms in which $G(t)$ is smaller may be neglected in the final synthesis. In the example given, the synthesis has been taken only to the $t = 12$ term for this reason.

The effect of these errors in $F(t)$ will be to make the values of F for large values of t more uncertain than those for t small. These erroneous values of F for large t will introduce spurious sinusoidal components of short wavelength into the final synthesis, so that local variations will be of little significance: the large-scale picture of $f(x)$ will be more or less correct, but the details will be unreliable.

In general a broad line gives Fourier components falling off more rapidly with increasing t than those for a narrow line. (This is illustrated by the values of $H(t)$ and $G(t)$ in tables 1 and 2.) We want $H(t)$ to fall off rapidly to zero compared with $G(t)$ so that we can neglect components which are divided by small values of $G(t)$ and are therefore subject to large errors. Therefore we want $g(x)$ to be narrow compared with $h(x)$. Thus we reach quite generally the conclusion (already noted by various writers in particular cases) that the standardizing lines, which give the "instrumental" broadening, must be as narrow as possible compared with the broadened lines of the specimen under examination, if we are to deduce reliable values for the corrected breadths and shapes of the lines.

§ 5. A CHECK ON THE NUMERICAL VALUES OF $f(x)$

The numerical values of $f(x)$ were checked by folding $f(x)$ and $g(x)$ numerically and comparing the result with $h(x)$. This may be done readily by taking the sum of products $f(y)g(x-y)$, as in equation (2); the values thus obtained were divided by 4800 to allow for the various factors introduced during the unfolding process and for the fact that the factor $1/a$ was omitted from equation (6) as

Table 5. Fold of $f(x)$ and $g(x)$, for comparison with $h(x)$

x	Fold	$h(x)$	x	Fold	$h(x)$	x	Fold	$h(x)$
-20	3	3	6	79	82	8	62	70
-19	5	4	-5	103	102	9	45	39
-18	6	5	-4	140	140	10	37	29
-17	7	7	-3	191	194	11	32	26
-16	9	9	-2	236	237	12	26	21
-15	12	11	-1	267	267	13	18	18
-14	14	13	0	272	277	14	16	16
-13	17	18	1	253	243	15	13	14
-12	21	22	2	227	233	16	9	12
-11	32	27	3	201	208	17	8	10
-10	38	33	4	177	177	18	3	8
-9	45	42	5	152	147	19	2	7
-8	51	50	6	120	125	20	1	7
-7	62	60	7	89	83			

explained in the text. Values of the fold thus obtained are given in table 5, with the original values of $h(x)$ for comparison. The values agree for the most part to within 2-3% of the largest value of $h(x)$; this is about the accuracy to be expected, since each of the several processes involved in unfolding, and the process of folding, have been carried out with an accuracy not exceeding about 1%.

§ 6. CONCLUSION

From the numerical example we see that, given a rapid means of summing Fourier series, the unfolding process can be applied to give a purely deductive method of obtaining corrected line shapes of x-ray diffraction lines. The same process might well have other applications, such as the deduction of corrected line widths in optical spectroscopy, and in statistical analysis of fluctuations in a case in which we know the frequency distribution of fluctuations due to a cause A acting alone, and due to causes A and B acting simultaneously, and wish to find the distribution for B alone (Whittaker and Robinson 1944 c).

ACKNOWLEDGMENTS

My thanks are due to Dr. H. Lipson and to Mr. M. A. Jaswon for several helpful discussions.

REFERENCES

- BIRKS, L. S., and FRIEDMAN, H., 1946, *J. Appl. Phys.*, **17**, 687.
DUMOND, J. W. M., 1924, Paper presented at the Pacific Coast Convention of the A.I.E.E., Pasadena, California, U.S.A., 13-17 Oct. 1924.
EDWARDS, O. S., LIPSON, H., and WILSON, A. J. C., 1942, *Proc. Roy. Soc. A*, **180**, 268, 277.
HÄGG, G., and LAURENT, T., 1946, *J. Sci. Instrum.*, **23**, 155. (References to several other Fourier-analysis machines given.)
JONES, F. W., 1938, *Proc. Roy. Soc. A*, **166**, 16.
KOCHENDÖRFER, A., 1937, *Z. Kristallogr.*, **97**, 469.
LIPSON H., and BEEVERS, C. A., 1936, *Proc. Phys. Soc.*, **48**, 772.
MEGAW, H. D., and STOKES, A. R., 1945, *J. Inst. Metals*, **12**, 279.
SHULL, C. G., 1946, *Phys. Rev.*, **70**, 679.
STOKES, A. R., and WILSON, A. J. C., 1942, *Proc. Camb. Phil. Soc.*, **38**, 313; 1944, *Ibid.*, **40**, 197.
WHITTAKER, E. T., and ROBINSON, G., 1944 a, *The Calculus of Observations* (London: Blackie, 4th edn.), p. 255; 1944 b, *Ibid.*, p. 280; 1944 c, *Ibid.*, p. 168.

LETTERS TO THE EDITOR

A Note on the ΔE Effect in Alnico

In this short communication we wish to draw attention to some experiments which have recently been made on the change in Young's modulus of a ferromagnetic accompanying magnetization, known as the ΔE effect. Up to the present, no results have been recorded for the high coercivity alloys, and in this note are given the results of an investigation of one particular permanent magnet material—alnico. The approximate percentage composition of the material was Al 10, Ni 18, Co 12, Cu 6, Fe 54, and specimens were prepared by casting in the form of rods of length 30 cm. and diameter 0.65 cm. Two such rods were available, one as cast or untreated, and the other heat-treated by maintaining it at a temperature of 1250° C. for 20 minutes, allowing it to cool at the rate of about 2° C. per sec., and then annealing it at 600° C. for about two hours, i.e. it was hardened and tempered. These rods were kindly supplied by Dr. K. Hoselitz.

The method used for the measurements, in which the resonant frequencies of longitudinal oscillation of the rods were observed, was the same as that already described by Street (1948). The rods were made to oscillate at their third harmonic at a frequency of approximately 25 kc/s. The maximum magnetic field intensity produced by the solenoid surrounding the specimen was 800 oersteds, and this gave intensities of magnetization of about 700 gauss in both the alnico rods.

It was found for both rods that the value of the resonant frequency when the specimen was magnetized to the maximum available intensity was only slightly greater than the corresponding value for the demagnetized material. Before the change in Young's modulus could be calculated it was necessary to determine the magnetostrictive change in the length of the rod. This was done by observing the system of Newton's rings set up between a plane glass disc attached to one end of the rod and a lens rigidly held near it. The value of the coefficient of magnetostriction for the available change in magnetization was found to be $+1 \times 10^{-6}$ approximately.

For both specimens the modulus increased steadily with intensity of magnetization and the maximum observed increase was 0.02%. From the form of the (I, H) curves it appeared that the specimens were reasonably close to technical saturation for the maximum applied field of 800 oersteds. Thus 0.02% represents a fair approximation to the maximum possible change in Young's modulus of the materials from the demagnetized to the saturated states. The corresponding figures for annealed nickel and cobalt specimens are 7.0 and 0.7% respectively. The values of Young's modulus for the two alnico specimens were equal within the limits of experimental error, the mean value being 17.0×10^{11} dyne/cm².

The decrements of longitudinal oscillations in the rods were determined from observations of the widths of the resonance curves obtained by variation of the frequency of the exciting current. The mean value of the decrement in each case was very small, 2.5×10^{-4} for the untreated specimen; only small variations about this mean were observed. In comparison, Ni and Co show large variations in decrement, the range for Ni being 10^{-3} to 5×10^{-2} and for Co 2×10^{-4} to 10^{-3} . There was some evidence to show that the mean decrement for the heat-treated rod was greater than that for the untreated rod. No great significance can be placed on the latter observation, however, since for such low decrements slight errors in the alignment of the rods within the apparatus may have appreciable influence on the result.

For any ferromagnetic specimen examined by the method used here there are certain ranges of the virgin curve and the hysteresis cycle over which the E.M.F. induced in the detecting coil is practically zero, and this behaviour was observed with the rods of alnico. Starting from the demagnetized state, no appreciable E.M.F. was induced until I was 100 gauss. For the hysteresis cycle the induced E.M.F. again approached zero over ranges of the intensity of magnetization approximately 200 gauss wide, around the coercive force points. Near these points, where the induced E.M.F. was small, the latter showed a pronounced time-lag in attaining a steady value following any change in the external field. No similar time variations of E.M.F. were observed over other regions of the magnetization curve, and the effect has not been noticed with the large number of other specimens already examined in this laboratory. The magnitude of the time-lag was larger for the untreated alnico specimen than for the heat-treated one. We have found by magnetometer methods that this variation of E.M.F. with time is associated with corresponding time changes in the intensity of magnetization.

It would appear from these results that rods of alnico, and possibly other high coercivity alloys, may be used with advantage in magnetostriction oscillator and filter circuits.

The University, Nottingham.
26th July, 1948.

R. STREET.
J. C. WOOLLEY.

THE PROCEEDINGS OF THE PHYSICAL SOCIETY

VOL. 61, PART 5

1 November 1948

No. 347

Sur les Conditions d'Auto-Oscillation des Systèmes Vibrants

PAR YVES ROCARD

École Normale Supérieure, Paris

A Discourse delivered 5 May 1948, on the occasion of the presentation of the Third Holweck Prize and Medal to Professor Y. Rocard

MS. received 5 July 1948

RÉSUMÉ. Le professeur Rocard a rassemblé en une doctrine plus compréhensible les conditions dans lesquelles les systèmes vibrants peuvent devenir auto-oscillants. On a depuis longtemps des moyens mathématiques très bien adaptés pour déterminer ces conditions. Il restait à les analyser et à en comprendre la nature. On peut sur quelques exemples voir la complexité du sujet ; l'auteur cite le roulement des bogies, des locomotives (la conicité des roues engendre l'auto-oscillation sous forme de "lacet"), des auto-oscillations du genre des coups de bélier entretenus dans les conduites, enfin les vibrations d'ailes d'avion, ou "wing flutter".

Il faut distinguer dans ces systèmes ceux qui résultent du couplage d'un certain nombre de circuits oscillants ou pendulaires ayant chacun leur période propre (le "wing flutter" est de cette sorte) et ceux qui résultent du couplage d'éléments n'ayant pas de période propre, mais seulement des constantes de temps : les télécommandes sont de cette sorte, et aussi les systèmes d'équations un peu arbitraires que l'on peut construire dans les théories économétriques. On peut établir alors les résultats suivants :

1. Si un système n'a qu'une fréquence propre, les forces qui peuvent le rendre auto-oscillant ont forcément le caractère d'une résistance *négative*.

2. Si un système résulte du couplage de plusieurs "degrés de liberté" ayant chacun leur fréquence propre, tout couplage passif (par self, capacité, inertie, etc.) *écarte* les fréquences propres composantes les unes des autres pour donner les fréquences propres du système, mais il existe des couplages *réactifs* qui peuvent rapprocher les dites fréquences propres, sans l'introduction d'aucune résistance positive ou négative. On démontre alors qu'un tel système devient auto-oscillant au moment où deux fréquences propres se confondent.

On peut imaginer dans les problèmes de "wing flutter" des cas très simples qui illustrent ce théorème, et on peut en donner des modèles électriques également très simples.

3. Dans un système tel que ci-dessus (§ 2) mettons maintenant des forces de résistance (proportionnelles aux vitesses) dans l'espoir d'amortir les auto-oscillations. On démontre ce résultat surprenant que, tant que les fréquences propres ne sont pas profondément modifiées, ces résistances ne font que faciliter l'auto-oscillation, qui prend naissance alors que deux fréquences propres se rapprochent mais ne sont pas encore confondues.

Ces résultats font d'ailleurs perdre à la théorie des oscillations de relaxation beaucoup de son importance philosophique, le mécanisme de la résistance négative n'apparaissant plus comme obligatoire dans l'auto-oscillation.

4. Dans le cas de systèmes tout à fait généraux, où on ne peut distinguer des fréquences propres composantes, il n'est naturellement pas possible de se livrer à une analyse aussi pénétrante, mais on démontre encore que si l'on attaque le système par une petite perturbation sinusoïdale dont on fait croître la fréquence, toutes les conditions de stabilité connues (Routh, Leonhard, Nyquist) expriment une vérité très simple : quand, à impédance croissante avec la fréquence, on passe par une résonance (la réactance s'annulant) le terme de résistance trouvé doit être positif.

JE dois tout d'abord exprimer mes remerciements à la Physical Society de Londres pour la haute distinction dont elle m'honore aujourd'hui. J'ai eu la rare chance, pendant la dernière guerre de pouvoir utiliser pour la défense de la Liberté mes propres possibilités scientifiques. Cette action comportait en elle-même sa récompense, et pour celle qui vient s'y ajouter aujourd'hui, je ne puis qu'en reporter le mérite sur ceux de nos collègues qui ont perdu la vie du fait de l'ennemi. Outre notre grand Holweck, je citerai spécialement MM. H. Abraham, E. Bloch, G. Bruhat. Tous trois ont dirigé le laboratoire de Physique de l'École Normale Supérieure, dont je suis maintenant chargé, et cette École s'honore aussi d'avoir vu beaucoup de ses élèves combattre pour la Liberté, aux côtés de ses meilleurs défenseurs, les Anglais. Elle sait de quelles traditions elle est la gardienne, et elle ne faillirait pas devant des tâches nouvelles qui se présenteraient.

Pendant la guerre, l'activité scientifique dans les pays occupés était profondément affectée par des circonstances en général fâcheuses, notamment une certaine asphyxie intellectuelle qui venait du manque de bibliographie en langue anglaise. Cependant certains savants, dont l'existence antérieure n'était peut-être pas assez repliée dans la méditation, ont trouvé dans les loisirs forcés de l'Europe Nouvelle l'occasion de mener à bien certaines tâches de longue haleine. Ainsi, plusieurs de mes illustres maîtres ont séjourné quelque trois semaines dans la prison de Fresnes. Ils ont assuré en être sorti les uns et les autres avec la solution enfin trouvée de problèmes mathématiques qui leur avaient paru jusque-là insolubles . . .

Bien que n'ayant pas été soumis à un traitement aussi efficace, c'est un peu de cette manière que mes propres travaux ont évolué, ce qui me fournit tout naturellement le sujet de cette "Address".

Dès 1934, j'ai eu à m'occuper de questions concernant la stabilité des systèmes mécaniques, et des conditions dans lesquelles ils deviennent auto-oscillants. Nous conviendrons donc aujourd'hui de rassembler sous le même vocable d'"oscillateur" des objets aussi différents que

- une locomotive qui a du lacet
- une lampe triode qui oscille
- une flute ou un tuyau d'orgue mis en mouvement par un jet d'air
- le wing-flutter des avions

etc... et nous chercherons ensemble quelles sont les causes qui provoquent l'auto-oscillation.

La locomotive est l'engin le plus difficile : on ne comprend même pas comment elle roule. Le Professeur Andrade est, je crois, l'auteur d'un petit livre sur les locomotives, il saura me redresser si mon exposé contient des erreurs :—Les roues des locomotives sont coniques, le rayon de roulement varie donc avec l'écart latéral de l'essieu sur la voie. Posons donc sur une voie un essieu avec ses deux roues, un peu de travers (figure 1), et poussons en avant, on voit sans trop de peine que l'essieu roule en avançant d'un mouvement sinusoïdal entretenu pur : l'oscillation est pendulaire, mais sa cause est purement cinématique. Si cependant l'on pose sur une voie un bogie avec ses deux essieux et ses quatre roues, les deux sinusoïdes qu'ont envie de décrire en roulant les deux essieux sont incompatibles avec les liaisons (les deux essieux doivent rester parallèles, etc...).

Un simple bogie ne peut donc même pas "rouler". Et cependant, il a bien l'air de rouler. Quel est donc son mouvement ? — On sait depuis Reynolds que

des déformations des rails et des roues autour de l'aire de contact finissent par rendre le mouvement possible moyennant l'intervention de certaines forces proportionnelles aux dites déformations. Ces forces ne dérivent pas d'un potentiel, elles peuvent soit amortir les oscillations latérales propres, soit les auto-entretenir: c'est généralement ce qui se passe au-dessus d'une certaine vitesse critique, et j'ai autrefois rempli deux petits volumes de calculs de telles vitesses critiques de locomotives. Je calculais très bien mais je comprenais très mal, lorsque les malheurs de la guerre obligèrent la généralité des Français à s'intéresser aux bicyclettes.

Or, on y trouve l'exemple le plus simple possible d'un oscillateur où la cause de l'auto-oscillation réside dans les liaisons non holonomes du roulement. Soit en effet le petit chariot représenté sur la figure 2, et largement simplifié: il est articulé en O, et toute la masse m est rassemblée en O. Un système de ressorts tend à mettre le bras l_1 dans la direction du bras l_2 . La masse m tend alors à osciller,

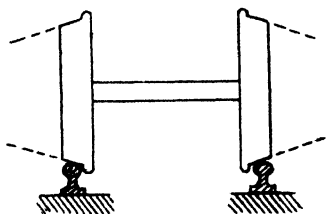


Figure 1.

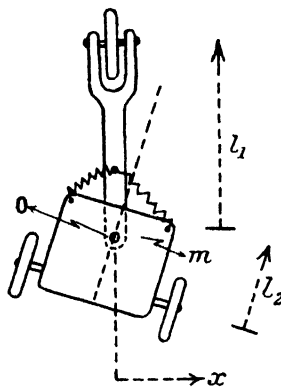


Figure 2.

en déplacement latéral, comme un pendule. Cette oscillation fait changer l'orientation des roues. Si alors on suppose que la machine se meut en avant ($V > 0$) ou en arrière, les liaisons du roulement réagissent sur le mouvement latéral de m , lequel est régi finalement par l'équation:

$$\left[\frac{d^2}{dt^2} + \left(\frac{1}{l_2} - \frac{1}{l_1} \right) V \frac{d}{dt} + \frac{K}{m} \left(\frac{1}{l_1} + \frac{1}{l_2} \right) - \frac{V^2}{l_1 l_2} \right] \frac{d^2 x}{dt^2} = 0, \quad \dots (1)$$

x étant l'abscisse latérale de m ; K est le coefficient du couple $K\psi$ exercé par le ressort lorsque les bras l_1 , l_2 font entre eux un angle ψ .

Cette équation est très instructive: (i) Pour $V = 0$ elle donne un mouvement pendulaire simple, sans amortissement. (ii) Si l'engin roule ($V \neq 0$) on a un terme d'amortissement, si le centre de gravité est en arrière du milieu quand on se tourne vers l'avant (donc, $l_2 < l_1$ si $V > 0$ et inversement $l_2 > l_1$ si $V < 0$). (iii) Dans le cas inverse, mettons $l_2 < l_1$ et $V < 0$ par exemple, on a un frottement négatif procurant un auto-entretien. (iv) Si la vitesse grandit encore on a, lorsque $V^2/l_1 l_2$ l'emporte sur $(K/m)(1/l_1 + 1/l_2)$ un nouveau type d'instabilité, avec écart exponentiel sans auto-oscillation. Si en particulier $l_1 = l_2$, l'amortissement positif ou négatif disparaît, et c'est seulement ce type d'instabilité qui intervient. Nous le retrouverons dans d'autres problèmes.

On voit donc que la locomotive était un exemple inutilement compliqué, et que des liaisons de roulement pur suffisent pour nous donner de beaux exemples d'oscillateurs.

Je ne m'étendrais pas davantage sur les oscillateurs à un seul degré de liberté, dont l'exemple le plus connu est celui de la triode oscillatrice, car en étudiant d'autres, (une aile d'avion vibrant en torsion seulement, une anche dans un tuyau d'orgue) nous retrouverions toujours ce résultat.

Un système à un degré de liberté devient auto-oscillant par une résistance négative.

Si l'on aborde maintenant les oscillateurs à plusieurs degrés de liberté les conclusions deviennent toutes différentes, et nous découvrons de nouvelles causes d'auto-oscillation. Quoique la théorie soit très générale, nous parlerons seulement pour deux degrés de liberté. Commençons par examiner les moyens mathématiques à notre disposition. En mettant en équation ces problèmes, nous trouverons généralement une équation différentielle portant sur une des variables, y , il sera suffisant de la considérer comme linéaire au moment où l'on sera à la limite de stabilité. Elle s'écrira donc symboliquement

$$a_0 p^n + a_1 p^{n-1} + \dots + a_n = 0, \quad \dots\dots (2)$$

p étant le symbole d/dt , et devenant aussi un paramètre algébrique lorsque l'on pose $y = e^{pt}$, et le système correspondant sera stable si l'équation (2) n'a que des racines à parties réelles négatives.

Les conditions d'une telle stabilité s'expriment par divers moyens, très connus, et en principe équivalents : le critérium de *Routh* (ou les déterminants de *Hurwitz*) ; le critérium de *Nyquist*.

Mais on fait ainsi des calculs aveugles, sans analyser les mécanismes physiques à la base.

Soit donc un système à deux degrés de liberté : sur la figure 3, on a représenté

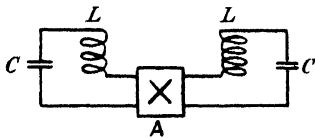


Figure 3.

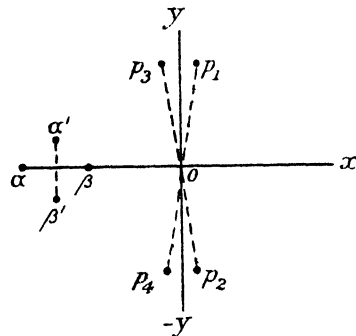


Figure 4.

deux circuits L, C et un organe de couplage A . L'étude de ce système nous mène à une équation $a_0 p^4 + a_1 p^3 + a_2 p^2 + a_3 p + a_4 = 0$, et les conditions de stabilité sont, d'après *Routh*, les suivantes :

$$\left. \begin{aligned} (a_2/a_0) - (a_3/a_1) &> 0 \text{ et en outre, } a_0, a_2, a_4 > 0, \\ (a_3/a_1) - \frac{a_4/a_0}{a_2/a_0 - a_3/a_1} &> 0 \text{ ainsi que } a_1, a_3 > 0. \end{aligned} \right\} \quad \dots\dots (3)$$

Elles sont bien classiques, mais on les écrit rarement sous cette forme. Nous avons voulu au contraire dans les équations (3) mettre en évidence que les termes

impairs a_3, a_1 n'intervenaient que par leur rapport. Or ces termes impairs viennent *uniquement* des résistances qui peuvent figurer dans les circuits, et si l'on convient de faire tendre les résistances vers zéro en leur gardant leurs valeurs relatives, les équations (3) montrent que les conditions de stabilité ne changent pas.

D'où la question suivante : *un système totalement dénué de résistances, positives ou négatives, peut-il être auto-oscillant ?*

Cela revient à étudier les racines de l'équation :

$$a_0 p^4 + a_2 p^2 + a_4 = 0. \quad \dots (4)$$

Sur la figure 4, nous avons représenté à la fois le plan complexe de la variable p^2 , et le plan complexe de la variable p .

α et β sont les deux racines en p^2 (deux nombres réels négatifs) lorsque $a_2^2 - 4a_0a_4 > 0$. Mais ces deux racines passent en α', β' lorsque $a_2^2 - 4a_0a_4 < 0$. Elles sont alors imaginaires conjuguées. Extrayant leurs racines carrées pour obtenir p_1, p_2, p_3, p_4 , les quatre racines de l'équation (4) on en trouve deux (p_1, p_2) à partie réelle positive, en divisant par deux l'angle argument de α' par exemple. Dans ce cas, l'équation (4) décrit un mouvement instable.

Or, le cas limite étant obtenu pour $a_2^2 - 4a_0a_4 = 0$, lorsque l'équation (4) avait ses deux périodes propres *confondues*. Nous avons donc obtenu la proposition suivante :

La propriété est visiblement générale, on peut la démontrer pour un nombre quelconque de degrés de liberté, et même pour une infinité (oscillations avec propagation, lignes électriques, tuyaux sonores...) et énoncer :

Dans un système à nombre quelconque de degrés de liberté, dépourvu de résistances positives ou négatives, l'instabilité naît avec la confusion de deux fréquences propres.

Si l'on cherche des exemples de tels systèmes, on trouve que tous les couplages *passifs* écartent les fréquences propres, donc ne peuvent donner naissance à l'instabilité. C'était bien nécessaire pour la conservation de l'énergie.

Ainsi, sur la figure 5, les deux pulsations propres sont (dans le cas simple où

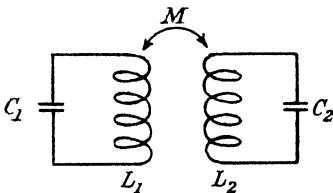


Figure 5.

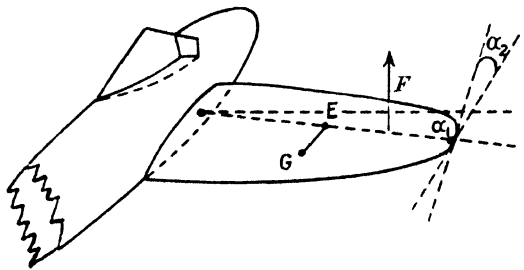


Figure 6.

$L_1 = L_2, C_1 = C_2$) sans couplage ($M=0$) $\omega_1 = \omega_2 = 1/\sqrt{LC}$, avec couplage ($M \neq 0$) $\omega_1 = 1/\sqrt{\{(L+M)C\}}$; $\omega_2 = 1/\sqrt{\{(L-M)C\}}$. Ces couplages passifs sont d'ailleurs caractérisés par la *symétrie* des coefficients dans les deux équations.

Un système mécanique équivalent au schéma de la figure 5 est fourni par *une aile d'avion* qui vibre en flexion-torsion (figure 6).

L'axe élastique E est le lieu des points où l'on peut mettre des poids sur l'aile sans la tordre. L'angle α_1 décrit la flexion de l'aile. L'angle α_2 décrit la torsion.

Si le centre de gravité G n'est pas sur l'axe élastique on conçoit sans peine qu'en empoignant l'aile en A et en la levant brusquement, la force d'inertie

appliquée en G va l'obliger à se tordre d'un certain angle α_2 : flexion et torsion sont donc *couplées* et, avec les variables choisies, le couplage est *par inertie*. On obtiendra donc finalement deux équations du type :

$$\left. \begin{aligned} \left(I_1 \frac{d^2}{dt^2} + K_1 \right) \alpha_1 + I_{12} \frac{d^2 \alpha_2}{dt^2} &= 0, \\ I_{12} \frac{d^2 \alpha_1}{dt^2} + \left(I_2 \frac{d^2}{dt^2} + K_2 \right) \alpha_2 &= 0. \end{aligned} \right\} \dots (5)$$

Par exemple, pour un avion de chasse, on peut avoir :

	Sans couplage ($I_{12}=0$)	Après couplage ($I_{12} \neq 0$)
Fréquence propre de flexion	13 cycles/seconde	10 cycles/seconde
Fréquence propre de torsion	24 cycles/seconde	30 cycles/seconde

Faisons maintenant voler l'avion qui possède cette aile, à une certaine vitesse V , elle va devenir un oscillateur, et d'ailleurs, cesser presque immédiatement : c'est là le "wing flutter", et voici comment on l'explique.

En vol, des forces aérodynamiques s'exercent, ces forces sont modifiées par la flexion et la torsion. Les nouvelles forces issues de la flexion sont négligeables (et ont d'ailleurs le caractère d'un amortissement), celles issues de la torsion sont proportionnelles à α_2 (angle qui vient augmenter l'incidence) et au carré de la vitesse. Elles peuvent être appliquées au voisinage de l'axe élastique, auquel cas elles ne réagissent pas sur la torsion elle-même, mais, en tout cas, elles exercent un moment de flexion considérable (force F sur la figure 6) de sorte que la première des équations (5) se réécrit ainsi (β , coefficient)

$$\left(I_1 \frac{d^2}{dt^2} + K_1 \right) \alpha_1 + I_{12} \frac{d^2 \alpha_2}{dt^2} = \beta V^2 \alpha_2, \dots (6)$$

donnant naissance au déterminant en p :

$$\begin{vmatrix} I_1 p^2 + K_1 & I_{12} p^2 - \beta V^2 \\ I_{12} p^2 & I_2 p^2 + K_2 \end{vmatrix} = 0.$$

On trouve sans peine que si le centre de gravité est en *arrière* de l'axe élastique ($I_{12} < 0$) il existe une vitesse critique, on s'en explique l'origine ainsi :

La seconde équation (5) montre que, aux fréquences ω basses ($K_2 > I_2 \omega^2$), α_1 est en *opposition de phase* avec α_2 ($I_{12} < 0$, $d^2/dt^2 = -\omega^2$). Dans l'équation (6), le terme $-I_{12} \omega^2 \alpha_2$ vient donc diminuer le rappel élastique $K_1 \alpha_1$ mais le terme $-\beta V^2 \alpha_2$ va l'augmenter, la fréquence propre de flexion va donc *augmenter* sans cesse avec la vitesse et se rapprocher de la fréquence de torsion : *l'avion meurt quand elles se confondent*.

On trouve facilement l'*équivalent électrique* d'un tel système : les équations (5) conviennent à la figure 5 si α_1 et α_2 représentent les *charges* des condensateurs C_1, C_2 . On réalise le cas de l'équation (6) par le schéma de la figure 7, où un amplificateur A attaqué par une tension en α_2 la transforme en une force électromotrice réinjectée dans le circuit de gauche (sous la forme d'un très fort courant passant dans une très petite résistance). Le gain de A est ici

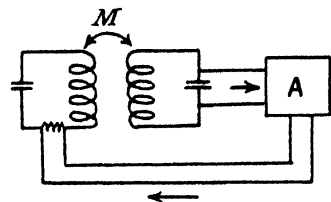


FIGURE 7

La question de l'auto-oscillation par confusion des fréquences propres est donc bien claire, elle exige d'ailleurs des forces qui ne dépendent pas d'un potentiel, mais qui n'ont pas à être en phase avec des vitesses, la résistance négative n'est pas le seul mécanisme qui amène à l'auto-oscillation, comme on avait tendance à le croire après les beaux travaux de M. van der Pol, continuant Sir Edward Appleton, sur les oscillations de relaxation.

Mais les systèmes réels sont beaucoup plus compliqués que ceux que nous venons de décrire, et il est nécessaire d'examiner l'effet des résistances sur la stabilité.

Introduisant alors des résistances dans un schéma tel que figure 7 nous trouvons ce très curieux théorème : à fréquences propres constantes, l'introduction de résistances positives ou négatives ne fait qu'augmenter la tendance à l'auto-oscillation :

S'il s'agit des circuits de la figure 7, l'amplificateur accrochera pour une valeur plus faible du gain ; s'il s'agit de l'aile de la figure 6 elle cassera pour une vitesse plus faible si sa structure est amortie.

Pour le voir, reprenons l'équation du quatrième degré, en p , et les conditions (3).

Posons $a_3/a_1 = \mu a_2/a_0$, une première condition de stabilité est $0 < \mu < 1$, quant à la seconde, elle s'écrit

$$\mu \frac{a_2}{a_0} - \frac{a_4/a_0}{(a_2/a_0)(1-\mu)} > 0 \quad \text{ou} \quad \frac{a_2^2}{a_0^2} > \frac{1}{\mu(1-\mu)} \frac{a_4}{a_0} \quad \dots\dots (7)$$

Or $1/\mu(1-\mu)$ est constamment supérieur à 4 quand μ varie de 0 à 1, donc la limite d'instabilité $a_2^2 = 4a_0a_4/\mu(1-\mu)$ est atteinte pour des valeurs des paramètres telles que $a_2^2 > 4a_0a_4$, c'est-à-dire avant que les deux fréquences propres, en train de se rapprocher, ne soient confondues.

On trouve sans grande peine des démonstrations tout à fait générales de cette propriété, pour un nombre quelconque de degrés de liberté. Cela revient à énoncer le théorème suivant :

Si un polynôme en x^2 , à coefficients tous positifs, a une racine double négative (donc quatre racines imaginaires pures pour x , valant $\pm j\omega$), et si on lui ajoute un polynôme quelconque ne contenant que des termes de degré impairs en x , le polynôme résultant aura au moins une racine à partie réelle positive.

On trouve aussi un théorème semblable pour les équations fonctionnelles (et non plus algébriques) auxquelles donne lieu l'étude de la stabilité des systèmes à propagation. Il y aurait beaucoup d'exemples à développer où l'on verrait comment la théorie précédente éclaire les conditions qui règnent à l'auto-oscillation, mais le cadre de cette conférence générale ne nous permet guère de nous étendre.

Consacrons plutôt les instants qui nous restent à déduire encore quelques conséquences nouvelles concernant le sens des conditions qui expriment la stabilité des systèmes.

Soit $H_{(p)}$ l'équation différentielle ou fonctionnelle qui exprime l'évolution de la variable y caractéristique du système $H_{(p)}$. $y = 0$. Si l'on pose $y = e^{j\omega t}$, on obtient

$$R + jX = 0, \quad \dots\dots (8)$$

R et X devenant deux fonctions de ω ; l'équation (8) en général n'est pas satisfaite, mais si exceptionnellement on peut trouver une valeur de ω telle que $R(\omega) = X(\omega) = 0$ c'est que les conditions concernant la limite de stabilité sont juste satisfaites.

Soit en particulier *un pendule amorti*

$$m(d^2y/dt^2) + f(dy/dt) + Ky = 0,$$

$$R = K - m\omega^2, \quad X = f\omega,$$

la limite de stabilité exigerait $f=0$, et si f était fonction de la fréquence, il faudrait plus particulièrement $f=0$ pour $\omega = \omega_1$ tel que $K - m\omega_1^2 = 0$. On remarque aussi que en faisant varier ω , lorsqu'on passe la *résonance* $K - m\omega^2 = 0$, on trouve une *résistance* $f\omega$ *positive*.

Il est à remarquer que l'expression $R/j\omega$, ici $K/j\omega + jm\omega$, est la *réactance* du système, et qu'elle est une *fonction croissante de la fréquence*. (Il y a un théorème qui l'exige.)

Or, il y a de fortes raisons de penser qu'un système quelconque, pour les fréquences voisines de sa fréquence de résonance est assimilable à un pendule avec une certaine masse, une certaine élasticité et un certain frottement (positif ou négatif):

En effet, au lieu de laisser le système au repos, attaquons-le par un signal $ae^{j\omega t}$, mesurons y en amplitude et en phase, de cette mesure résultent $R(\omega)$ et $X(\omega)$. Attachons-nous à une valeur ω_1 de ω telle que $R(\omega)_1 = 0$, qui donne $X(\omega_1) = X_1 = f_1\omega_1$, essayons les fréquences voisines $\omega_1 - \Delta\omega$ donne $R(\omega_1 - \Delta\omega) = -\Delta\omega(dR/d\omega)$ (mesuré). Essayons de poser $R = K_1 - m_1\omega^2$ pour ω voisin de ω_1 on aura

$$(1) \quad K_1 = m_1\omega_1^2,$$

$$(2) \quad -2m_1\omega_1\Delta\omega = -\Delta\omega(dR/d\omega)_1 \text{ (connu),}$$

d'où

$$m_1\omega_1 = -\frac{1}{2}(dR/d\omega)_1 \text{ et } K_1 = -\frac{1}{2}\omega_1(dR/d\omega)_1.$$

On trouve donc K_1 et m_1 . Si ces deux termes ont des valeurs *positives*, il est clair que, physiquement, si le système est stable, c'est que le frottement f_1 doit être positif, donc X_1 doit être trouvé positif. Or K_1 et m_1 sont positifs si $(dR/d\omega)_1$ est négatif, donc la condition de stabilité que nous trouvons est la suivante:

L'expression $X(dR/d\omega)$ doit être négative pour toute valeur $\omega_1, \omega_2, \dots$ de ω qui annule R comme R ne dépend que de ω^2 , on peut, plus simplement encore dire:

$$X(dR/d\omega^2) < 0. \quad \dots\dots(9)$$

En termes physiques, on dirait plutôt: la condition de stabilité est qu'on trouve un frottement X positif à toute résonance de réactance $R=0$ *parcouru dans le sens des impédances croissantes* (sinon, on trouve une masse négative, une élasticité négative, et la stabilité veut alors un frottement négatif, tous les signes de l'équation étant alors changés).

Les remarques qui précèdent, *extrêmement simples*, nous donnent alors la clé des conditions mathématiques déjà connues, mais dont on comprend mieux maintenant la nature.

Ces conditions connues, pour la stabilité, sont (i) le critérium de Nyquist, (ii) l'entrelacement des racines de R et de X . Il a déjà été démontré que ces conditions sont équivalentes. Or notre condition (9) est équivalente à l'entrelacement: En effet à chaque racine successive ω_1, ω_2 de R , $dR/d\omega^2$ change de signe, donc, dans le cas de la stabilité, $X(\omega_1), X(\omega_2)$ etc. . . . sont alternativement de signes contraires, donc les racines de X et de R sont entrelacées.

Comme exemple de problèmes de stabilité traités suivant ces méthodes, nous pourrions citer beaucoup de problèmes de télécommande, mais sans doute la

plupart des personnes qui m'écoutent prendront-elles plus d'intérêt à la solution de quelques problèmes concernant la stabilité des automobiles.

J'en citerai rapidement deux :

1. Stabilité de route d'une voiture qui roule, direction bloquée.

La figure 8 représente l'avant de cette voiture, vu de dessus, roulant normalement. Les aires de contact des bandages avec le sol sont hachurées. La figure 9 représente ce qui se passe si une force latérale agit sur la voiture, les roues se déforment, sans que l'aire de contact change de place, mais elles vont du fait de

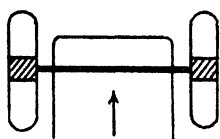


Figure 8.

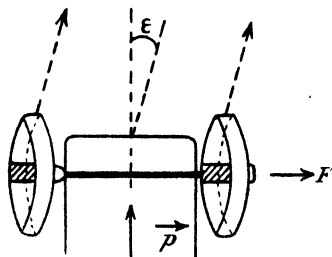


Figure 9.

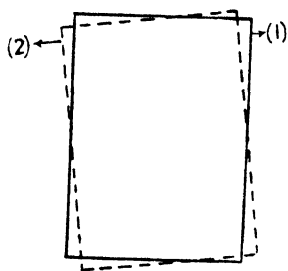


Figure 10.

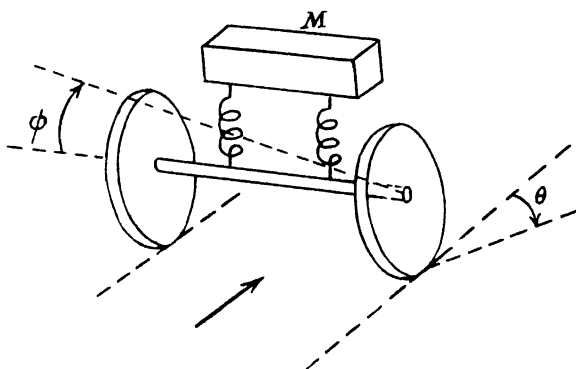


Figure 11.

leur déformation rouler dans une direction oblique. L'angle ϵ est proportionnel à la force F .

La prise en considération de cet effet permet la mise en équation du "roulement" compte tenu des oscillations du rectangle (1) représentant la voiture par rapport au rectangle (2) des aires de contact, ce rectangle (2) étant lui-même susceptible de "rouler" dans une direction oblique suivant les forces qui tirent sur lui, et qui viennent du déplacement relatif du rectangle (1) (figure 10).

La résolution de ce problème m'a conduit aux résultats suivants

- (a) Si les quatre pneus sont également gonflés, la voiture ne présente pas de "vitesse critique d'instabilité" si le centre de gravité tombe entre le milieu et un certain point un peu sur l'avant.
- (b) Si le centre de gravité est, pour simplifier, pris au milieu de la voiture, celle-ci ne présente pas de vitesse critique d'instabilité si les pneus avant sont un peu moins gonflés que les pneus arrière, dans une proportion d'ailleurs limitée.

Telles sont les meilleures conditions de stabilité d'une automobile. J'espère que la prospérité bientôt retrouvée dans nos deux pays nous permettra à tous de les expérimenter

2. *Le Shimmy des voitures comportant un essieu avant.*

Ce vieux phénomène, qu'on tend à oublier, était bien gênant en 1928 environ. En 1948, il reste intéressant, parce que c'est un cas pur d'oscillateur selon nos idées, avec l'auto-oscillation par confusion de deux fréquences propres. Le figure 11 montre un essieu avant, avec deux roues, suspendu sous une voiture. Les masses et les ressorts de la suspension, des pneus, donnent une période propre d'oscillation pour l'angle ϕ . Mais les roues roulent vers l'avant, et vont créer un effet gyroscopique. Si par un mouvement ϕ , la roue de gauche s'élève, le plan des deux roues change d'orientation, l'effet gyroscopique le fait tourner de θ , la voiture roule en cercle, de *part les liaisons du roulement*. Il naît alors une *force centrifuge* F qui pousse la masse M vers l'extérieur et tend à donner une variation contraire à ϕ ; voilà le couplage dissymétrique qui rend le système auto-oscillant: Une *rotation* θ produit un *couple* sur ϕ (en plus de l'effet gyroscopique avec son *couple* en $d\theta/dt$ agissant aussi sur ϕ), tandis qu'une rotation ϕ ne produit sur ϕ que le couple gyroscopique.

Le nouveau couple introduit est proportionnel à V^2 , on voit sans peine qu'il confond les fréquences propres en θ , ϕ à une certaine vitesse critique V_c , dont on obtient l'expression.

Si l'on réduit le Shimmy à ces remarques très simples en le schématisant un peu, on y trouve *une seule cause* d'instabilité, et on a un schéma tout à fait calqué sur celui de l'aile d'avion de la figure 6.

CONCLUSION

Sur les exemples développés au cours de cette lecture, nous espérons avoir montré qu'il était possible de se livrer à une véritable analyse des causes des auto-oscillations, au lieu de se borner à une discussion mathématique rigoureuse moins aveugle de leur stabilité. On trouvera une doctrine complète exposée à ce sujet dans l'ouvrage "Dynamique générale des vibrations" paru en France en 1943 et dont une seconde édition plus étendue s'imprime en ce moment. Une traduction anglaise est en cours, sans qu'on puisse encore préciser sa date de parution.

Je tiens en terminant à adresser un vif remerciement à la Physical Society de Londres pour m'avoir donné cette occasion très flatteuse pour moi d'exposer ces quelques travaux.

The Growth of Crystals

By G. P. THOMSON

Imperial College, South Kensington, London

Thirty-second Guthrie Lecture, delivered on 4 June 1948

IN choosing a subject for the Guthrie Lecture it seems important to consider the use to which such a lecture can be put. There is so much published now on physics that any addition to the literature needs justification. I have no great discovery to report, but there is a field in which considerable work has been done at Imperial College both in the Physics Department and in that of our President *, and which I think it may be profitable to bring to your notice because it seems very promising, and it is one which has not perhaps received due attention in this country—the growth of crystals.

While the last thirty years have seen an enormous increase in knowledge of how the atoms are arranged inside crystals, there has not been a corresponding advance in understanding of the kinetics of the process of growth, or of the causes which determine outward form.

Certain thermodynamical arguments have long—perhaps too long—dominated this study, and though I hope to show that they are liable to misuse, it is best to start from them.

In general a crystal must start from small beginnings, and the same sort of argument which governs the growth of water drops in a mist or a Wilson chamber applies here. Free energy resides in the surface separating two phases and retards the formation of a second phase inside the first. On dimensional grounds a surface effect must be more and more important compared with a volume effect the smaller the volume. Hence for very small drops the effect of surface tension is greatly to increase the vapour pressure with which the drop is in equilibrium, according to the well known formula $\ln p_r/p_x = 2v\sigma/rkT = \delta$, where σ is the surface tension, r the radius of the drop, v the volume occupied per molecule of the liquid, and δ is the supersaturation measured as the logarithm of the ratio of the vapour pressure in equilibrium with the drop to that in equilibrium with a flat surface. The drop is of course unstable. Increase of radius makes the vapour oversaturated with respect to it so that it grows more.

If a crystal is started without a previous nucleus we should expect a similar expression to apply. It is true that Gibbs (1906)† pointed out—“..... We cannot imagine a body of the internal structure and external form of a crystal to grow or dissolve by an entirely continuous process, or by a process in the same sense continuous as condensation or evaporation between a fluid and a gas, or corresponding process between an amorphous solid and a fluid. The process is rather to be regarded as periodic.” Kossel (1934) has raised a similar objection. Yet by considering the growth of a complete layer over the surface of a small crystal we can derive an expression similar in form to the above with the modification that the linear parameter corresponding to r can only increase by discrete steps corresponding to the addition of layers one cell thick all over the crystal, while the

* Professor G. I. Finch.

† I am indebted to Mr. Humphreys-Owen for calling my attention to this and the following quotation.

quantity corresponding to σ may have to be averaged for different faces if the crystal includes more than one "form".

Stranski (1928) has shown that for a cube of side $2r$,

$$\ln \frac{p_r}{p_\infty} = \frac{2v}{kT} \left(\frac{\sigma}{r} + \frac{K}{2r^2} \right),$$

where the second term is due to edge energy.

We may digress here to note how, from the thermodynamic point of view, the equilibrium shape is determined. As P. Curie long ago pointed out, the crystal truly in equilibrium with solution or vapour will take a shape such that the free energy is a minimum for a given volume. Take some point P, inside the crystal. Let p_1 etc. be the perpendiculars from P on the faces of the crystal of area F_1 etc. Then the volume V is the sum of the pyramids with P as vertex and the faces as bases: $V = \frac{1}{3} \sum_1^n p_1 F_1$, but the free energy $A = \sum_1^n \sigma_1 F_1$, where σ_1 etc. are the surface tensions. Now let one set of faces grow at the expense of another. The change in volume can be written either as $\sum_1^n F dp$ or $\frac{1}{3} \sum_1^n (F dp + p dF)$. If we keep the volume constant both are zero, and hence $\sum_1^n p dF = 0$. But the condition that A is constant

gives $\sum_1^n \sigma dF = 0$ (neglecting any variation of σ with F), so that faces for which it is possible to vary F independently must be so situated that $p \propto \sigma$. This implies that the point is taken midway between each pair of faces with equal σ , and when so taken p/σ is constant for all possible forms. This result is due to Wulff (1901).

But all this has its limitations. When we consider quantities, we find that for a crystal of a few cubic centimetres the difference in the equilibrium concentration between one form and another, e.g. (100) and (110) for rocksalt (Brandes 1927) is only of the order of 1 in 10^6 . (Notice that, like all surface energy effects, it vanishes as the crystal grows indefinitely). This is too small to make any appreciable difference except perhaps for the slowest possible growth such as might occur in geological time. In any laboratory process small fluctuations of temperature or of concentration due to diffusion will far outweigh this.

What matters is the *rate* of growth, and this may well be governed by some initial step. Gibbs (1906) early recognized this: "Since the difficulty in the formation of a new layer is at or near the commencement of the formation, the necessary value of the Potential may be independent of the area of the side except when the side is very small. The value which is necessary for the growth of a crystal will however be different for different kinds of surfaces, and will probably generally be greatest for the surfaces for which the Surface Tension is least. . . . As a final result, a large crystal will generally be bounded by those surfaces on which the deposit of new matter takes place least readily. If one kind of surface cannot form a closed figure, a crystal will be bounded by two or three kinds of surfaces determined by the same condition. The kinds of surfaces thus determined will probably be generally those for which the Surface Tension has the least values. But the relative development of different kinds of sides will not be such as to make the sum of the Surface Energies a minimum." Gibbs's conception has been developed by Kossel (1928) and Stranski (1943) and their pupils in the following way. Consider a simple cube lattice built up of cubes of matter and suppose for

simplicity that the only forces are those between the faces of the little cubes that come into actual contact.

Suppose the crystal has grown till its side is n units and that it is a cube. Then of the $6n^3$ faces all will be in contact except the $6n^2$ on the surface. If ϕ is the binding energy for each pair of faces, the average binding energy per unit is $(3n^3 - 3n^2)\phi/n^3 = 3\phi(1 - 1/n)$. When the crystal is infinite this is 3ϕ , and a unit which requires this work to detach it will just be in equilibrium with a solution or vapour saturated for a large crystal.

Now build a new layer on one face. The first unit, say at a corner, only has one face in contact with the crystal—work ϕ . The next along an edge has work 2ϕ , there are $2(n-1)$ of these or $2(m-1)$ if the layer only covers m^2 of the units, $m < n$. The remainder all have the full work 3ϕ . This corresponds to what Kossel calls the “repeatable step”. Once we have got over the initial difficulty, the crystal will grow or diminish according as the solution is under- or over-saturated for this “repeatable step”. But what is the measure

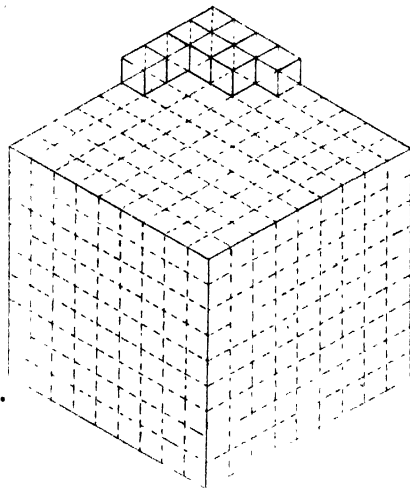


Figure 1.

of the initial difficulty? The total separation energy for the whole plane is $\phi + 4(m-1)\phi + (m-1)^2 3\phi = 3\phi m^2 - 2\phi m$. For a line added to an incomplete plane the separation energy is $3\phi m - \phi$. The defect is due to the fact that we have extended two faces of the cube by an amount $r^2 m$ each, where r is the side of a unit cube, and added two new steps of equal area, and so produced surface energy $4mr^2\sigma$; so $4mr^2\sigma = 2m\phi$ or $2\sigma = \phi/r^2$, as is otherwise obvious.

The defect in separation energy requires an extra concentration of energy in the form of a probability fluctuation. The amount required is $2m\phi = 4mr^2\sigma$, and increases in proportion to the side of the embryo produced. Small fluctuations are frequent, but unless the embryo produced is large enough to be in equilibrium with the degree of supersaturation present, it will relapse back into its elements and come to nothing. The production of an embryo thus becomes more likely the greater the supersaturation, because the greater this is the smaller the embryo needs to be.

While the work for the “repeatable step” is the same for a small or large embryo the small embryo is unlikely to survive. It is, indeed, equally likely to gain or lose at each step, but its chance of growing really large is small, for it will probably be extinguished by chance fluctuations. The case is rather analogous to a man gambling at Monte Carlo. If he goes on he is pretty sure to lose all his capital, even apart from the zero in favour of the bank, just because of probability fluctuations. The smaller energy of attachment of the first few pieces acts against the crystal something like the zero against the gambler. The crystal can only be saved by the solution being supersaturated. There appears to be no way of supersaturating Monte Carlo!

A similar explanation may apply to the growth of large crystals at the expense of small, as in the process of annealing.

In calculating the energy fluctuation necessary to produce an embryo which will be in equilibrium with a particular degree of supersaturation account has to be taken of the work done by the gaseous (or osmotic) pressure, and this reduces the fluctuation required from $F\sigma$ to $F\sigma/3$ in the case of three dimensions (Volmer 1921) and from $u\rho$ to $u\rho/2$ in the case of two dimensions (Volmer and Adhikari 1925, 1926), where u is the perimeter in equilibrium with the supersaturation and ρ is the work of separation per cm^2 surface.

Now the trouble is that for moderate supersaturations the energy fluctuation is extremely large. Thus Brandes (1927) calculates for rocksalt with 1% supersaturation a value of 2.3×10^{-9} erg for the surface embryo. This is $6 \times 10^4 kT$ at room temperature and growth should be negligible, while in fact a crystal of visible size would grow fairly fast at such a supersaturation. The chance of energy concentration includes of course the Boltzmann factor, $\exp(-W/kT)$.

Amsler (1942) has made experiments on the delay times for the growth of crystals from solution, detecting the onset of crystallization by measuring the change in the electrical conductivity. The delay time is subject to a probability fluctuation, but when the stirring is good the results are surprisingly consistent. The chance should go as $\exp(-C/\delta^2)$ where $C = 16\pi M^2 N \sigma^3 / 3\rho^2 R^2 T^3$, where M is molecular weight, ρ is the density, N is Avogadro's constant and δ the supersaturation. The variation with δ is as expected, but the constant gives a value of σ of the order of 1 dyne/cm. Now σ is the difference of the surface tensions of the solid and liquid if the contact angle is zero. The calculated value for rocksalt is 107.5 dyne/cm. for a large crystal and that for the solution is 78.8 dyne/cm. giving a difference of about 30 dynes which is cubed in the exponential so that the factor here is of the order 30^3 . Amsler thinks that the change in surface tension of the solid with size may account for the difference but it seems unlikely that it would so closely approach that of the liquid, especially as results with NaCl and KCl are closely similar.

In any case, the difficulty remains in the two-dimensional case which Amsler's work merely confirms.

The way out seems to lie in an early suggestion of Volmer (1921), to whom indeed much of the above theory is due. He supposed that the surface of a growing crystal is covered with a mobile supersaturated layer. He and Estermann (1921) brought forward evidence for the motion of mercury atoms over the surface of a growing crystal, showing that the material reaching a face directly from the vapour might be many times less than that which actually formed on it. He also (Volmer and Weber 1926) found that crystals of benzophenone appeared to lose and gain matter by flow over the surface, and measured the diffusion of a very thin film of benzophenone over a glass plate. It should be mentioned perhaps that the transference by vapour pressure was appreciable in these latter experiments. Other instances of atoms migrating over surfaces are the effects observed by Stern, Estermann and Cockcroft with molecular beams, and the technically important effects studied by Langmuir and others in which alkali atoms migrate over tungsten.

If now we suppose that a layer exists on a growing crystal which can diffuse rapidly over the surface, but which is much more concentrated than the vapour or solution round the crystal, we can explain the rapid formation of embryos in such a case. The fluctuations of energy needed to form an embryo of given size, and the size of the embryo needed to produce growth, are now quite different from what we have calculated above. The equilibrium condition is enormously

hastened, as by a catalyst. There is strong evidence for the existence of such a layer, not only in the work of Volmer, but in a line of experiments started by Berg (1938) and recently developed at Imperial College by Humphreys-Owen, to whom I am indebted for allowing me to give some of his results. The experiments consist in illuminating a crystal growing from solution between optical flats in such a way as to produce interference fringes. From these the variation of concentration from point to point of the solution can be determined, lines of equal concentration can be drawn and the flow normal to the force of the crystal determined, since this by the law of diffusion is proportional to the normal concentration gradient. Berg and Humphreys-Owen agree in finding that the surface of the growing crystal is far from being a surface of constant concentration (see figure 2),

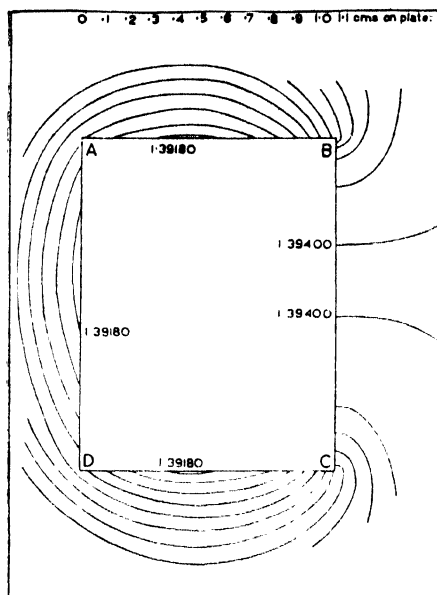


Figure 2. The lines show surfaces of equal concentration. The face at the right has stopped growing, others are growing normally. The concentration at the corners is greater than that in the middle of the face.

and also that the normal flow is not constant along a growing face, being most at the centre and least at the corners. Since the face in fact grows flat, there must be a flow of material from the centre outwards in a layer too thin to be observable, which means less than about 0.01 mm. The distribution of flow is sufficiently often symmetrical about the middle point for one to regard this as the normal case, but there are frequent cases of exception due presumably to greater ease of growth at one side or the other (figure 3). Humphreys-Owen finds that the maximum flow in the layer is proportional to the inflow in the normal case, and is about 7% of it.

At this point I should like to show you a film kindly lent me by Mr. Ferguson of Imperial Chemical Industries which shows some experiments on crystal growth: the film has been shown before but the results have not yet been published. You will notice that the growing face increases in thickness discontinuously. Steps

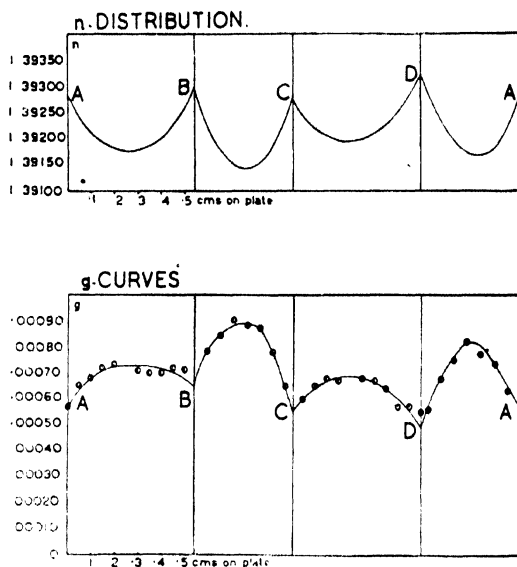


Figure 3. The top half shows the distribution of concentration over the four faces AB—BC—CD—DA. The lower half shows the normal flow at various points on these faces. The crystal to which this figure refers, unlike that in figure 2, is growing normally on all faces.

are formed, often at a corner, which sweep across the crystal. Marcellin (1918) seems to have been the first to observe this phenomenon. Sometimes one step will catch up on another and the whole move forward together. I suggest that we see here the absorption of the mobile layer at the step, and that the speed with which the step advances is governed at least in part by the mobility of the layer. Thus in figure 4 material from the old face of the crystal will be flowing towards the advancing step to build it up, while a new layer will be reforming on top of the step and flowing in the opposite direction. Both these layers will flow in a way determined by some law of diffusion, though it seems reasonable to suppose that their thickness is limited to some saturation value. If the step is shallow it will be able to advance more rapidly than if it is deep, for it will not have to wait for so much material to flow in from parts of the mobile layer ahead of it. This flow can only occur by a species of surface diffusion, depending on changes in the concentration of the mobile layer. It ought to be possible from Humphreys-Owen's results to deduce a good deal about the properties of the layer, such as its coefficient of diffusion. Experiments of this kind should be combined with experiments of the Ferguson type under comparable conditions.

Kowarski (1935), who was one of the first to observe the stepwise growth of crystals—in his case paratolindine from the vapour—observed a curious effect, which he attributes, I think rightly, to the Volmer layer. He found that, by adjustment of temperature, liquid drops could be formed on the crystal surface. In certain cases when the crystal was growing in area—the crystals were in the form of thin flakes—these drops were drawn to the advancing edge of the crystal, and especially to the foremost corner, where they were held as though by an elastic force. This seems to indicate a flow of the mobile layer to the edge of the face which it is prolonging, carrying the drop with it. All this fits well with the Volmer picture. Kowarski also studied the rate of surface growth in different directions and showed that it was inexplicable unless the vapour could move over the surface from where it first struck. He finds a maximum velocity for the layer of the order 2 or 3×10^{-2} cm/sec.

In considering crystal growth one's interest naturally centres on the atomic processes, but there are other factors which may be more important in practical cases—for example diffusion. If a crystal is growing fast it will exhaust the material in its immediate neighbourhood, and any protrusion will be more favourably situated than the rest. We have now seen how up to a point a crystal can continue to grow flat faces though different parts of them are at different concentrations, but there must be a limit to this, and when it comes the crystal grows out to unexploited regions. This leads to dendritic growth. Even for a crystal of compact shape the effect is important and may govern the rate of growth. We can make a crude model (Valeton 1923, 1924) which brings out the main points, by supposing that there is a discontinuity of concentration $C_s' - C_s$ at the surface of the crystal, supposed a sphere of radius a , and that the flow of material into the crystal is $k(C_s' - C_s)$ per cm^2 . Then if D be the coefficient of diffusion (taking 1 cm^3 of solid crystallite as unit of material) it is easy to show that

$$\frac{da}{dt} = \frac{C_\infty - C_s}{a/D + 1/k},$$

where C_∞ is the concentration of the distant solution. The two terms in the denominator can be regarded as series resistances, one due to diffusion in the bulk

of the liquid and the other due to the surface layer. If the resistance due to diffusion is the important one, da^2/dt is constant, but if that due to the surface layer is predominant, then da/dt is constant.

Humphreys-Owen finds that for his crystals the first was the case (his crystals were in two dimensions which gives a somewhat different law, but the difference is slight). It should be noticed, however, that the relative importance of the terms depends on a , but his crystals were fairly small, approximately 2 mm., and the law should hold *a fortiori* for larger ones.

It is well known that the habit of crystals is influenced by the presence of colloids in the solution, which favour dendritic growth. They probably act by

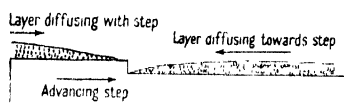


Figure 4.

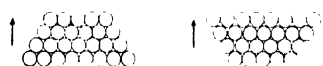


Figure 6.

(a)

(b)

(a) Quick-growing face is extinguished.

(b) Quick-growing face increases.

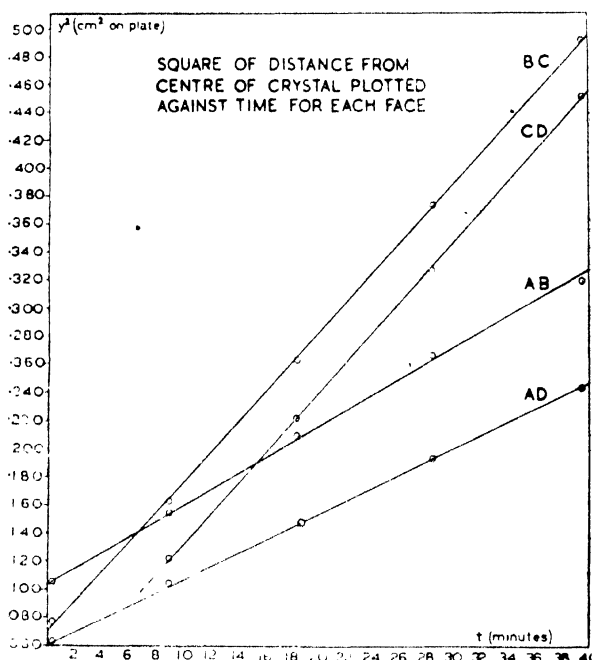


Figure 5. Showing normal growth of a crystal for which the square of the characteristic dimension increases linearly with time. The zero of time is the moment when observation started, not that when the crystal started growth.

reducing the coefficient of diffusion of the salt and so increasing the difficulty of supply of the material, making this relatively more important than the resistance to embryo formation.

There may, however, sometimes be another effect which, indeed, almost certainly occurs when some dyes are present in solution. These are preferentially absorbed on certain faces, and so inhibit their rate of growth. It is usually supposed that the faces finally produced are those which grow most slowly, the more quickly growing faces extinguishing themselves. This is not self evident, as is seen by a comparison of figure 6 (a) and (b). It is conceivable that in certain circumstances a face might extend as it grows as in (b). This may occur in certain cases of very rapid growth where the supply of material is the determining factor to the exclusion of considerations of energy.

Another way in which non-atomic considerations may modify crystal growth is the purely geometrical effect of surroundings. Thus crystals which tend to form flat plates and are grown from vapour on a solid substratum tend to be orientated with the plates parallel to the substrate (Thomson and Cochrane 1938). If the vapour is directed as a molecular jet they tend to be orientated normal to the jet, and when both causes are present a compromise is reached. These effects may be explained by the greater ease of growth afforded to a plate which happens to start parallel to the solid surface, or in the second case by the greater mass intercepted by the face normal to the stream. In both cases we must suppose that the molecules that strike the surface form a mobile layer over it which moves so as to build up the edge of the discs, as in Volmer and Estermann's experiment.

Professor Finch has informed me of a similar effect when crystals are grown in fine scratches.

It is pretty general experience that very thin films formed by spluttering or evaporation are, if not amorphous, at least micro-crystalline (Thomson and Cochrane 1938). The work of Andrade and Martindale (1935) on gold is a good example of this, as is also Lotmer's (1945) on antimony. Increased deposition or heating leads to crystallization. Andrade and Martindale found that the first stage was to form spherulites.

When growth occurs it seems to do so over the surface of the substrate, and the crystals show strong orientation with regard to this surface (Thomson and Cochrane 1938). It has been reported (Kowarski 1935) that one crystal can sometimes influence the orientation of another at a distance, which suggests a very thin mobile crystalline layer. The evidence for this is not yet conclusive, but it would be intriguing if the Volmer layer turned out to have an ordered structure of some kind. Other cases of crystals increasing in size with time are probably examples of the tendency for the large to swallow up the small.

So far we have considered a crystal growing from a seed of its own composition, or from none at all. Very interesting effects occur when one crystal is grown upon another. The classical case is the growth of sodium nitrate on calcite. The two are isomorphous and the sodium nitrate grows from solution with its axes parallel to those of calcite. Many other cases are known, for growth both from solution and from melts, for electrolytically deposited crystals and for crystals grown from vapour.

Isomorphism is not essential. Sodium nitrate will form orientated crystals from the melt on mica (Lassen 1934), and very large crystals can be grown in this way, but though mica is monoclinic, the top layer of atoms is pseudo-hexagonal and similar to the (111) face of sodium nitrate.

The view is often expressed that orientated growth of this kind, or epitaxis as it is called, depends essentially on a near equality of spacing between atoms in the top layer of the substrate and those in the layer of the crystal that grows to join it. Certainly such an agreement exists in many of the most striking cases, but I would like to suggest that the importance of this has been exaggerated and that other considerations are equally important. In particular I think that the existence of steps and cracks in the mother crystal are often the determining factor. Atoms align themselves along these steps, which become a zone axis of the daughter crystal, some prominent face grows out over the surface of the mother crystal, and thus orientation is established. Certainly this is more easily done if there is a similarity in the atomic patterns of the two layers, but even this is not essential.

Andrade and Martindale have shown the importance of cracks in determining the siting of crystals of silver and gold grown on glass and diamonds. Strictly, this is not quite the same thing, since they were concerned with the way in which minute crystals scattered over a substrate surface grouped themselves while we are interested in the direction of the axes of the individual crystals. Still, the attraction of the cracks for the atoms that are trying to form up as a crystal is strikingly shown in their work.

I should like to call your attention to three or four examples in which epitaxis has been observed for groups of crystals.

Let us take first the classical work of Royer (1928) on the growth of alkali halides on mica. He found that among 25 alkali halides, both of the face- and body-centred types, 11 showed good orientation on mica. In all cases the (111) face of the halide coincided with the cleavage of the mica. The pattern on this face is of course one of equilateral triangles, and in all cases but one these were so turned as to fit the corresponding pattern of the mica. If the difference in the dimensions of the pattern was less than 12% orientation occurred, if more it did not. In one case, KCl, orientation sometimes occurred with a 14.5% discrepancy. But there is another type of orientation in which the triangles are set at right angles. The fit here is much better for KCl—almost perfect—but it applies only to $\frac{1}{4}$ of the atoms. It is the less common form of orientation of the two. Four other of the halides could adopt this form of orientation with less percentage discrepancy than in the one they actually chose, but this is not observed. This goes to show that the specific nature of the planes and lines on which the fit occurs is more important than the actual numerical closeness of agreement.

Another example is given by metals deposited by spluttering on rocksalt. They have been studied by Lassen and Brück (Brück and Lassen 1935, Brück 1936) using the method of electron diffraction. They find many examples where the orientation actually chosen is not that which gives the closest fit, as can be seen from table 1. Brück interpreted his results in terms of the mean distance of atom in the mother and daughter crystals averaged over a cell, which he found tended to be a minimum, but in view of the different electrical character of the Na and Cl atoms in the surface it seems doubtful if much weight can be attached to this.

In all Brück's cases important lines of atoms coincided in the two crystals, and this appears also from the experiments of Rüdiger (1937), who studied the deposit of Au, Ag, Pt on calcite, mica and fluorite.

It should however be pointed out that in Brück's case, at least, there is no single direction on the mother crystal, which is always a zone axis for the daughter.

When metals are deposited on metals there is a strong tendency to follow completely the structure of the underlying metal as shown in the work of Cochrane (1936) and Finch and Sun (1936), who in both cases deposited the metal electrolytically and tested the result by electron diffraction. Many suitable metals have nearly the same structure and spacing, so on almost any theory one would expect them to fit in this way. Interesting exceptions are iron on palladium and gold, where Finch and Sun found cube planes in contact, but the face diagonal of the iron parallel to the cube edge of the substrate. This gives a good fit. The same authors found a peculiar effect of nickel on gold in which the same arrangement as that for iron occurs, though complete parallelism would actually give a *better* fit.

Cochrane found that a substrate of copper forces nickel to adopt the copper spacing, which is 2.5% different, for very thin deposits. The strain causes

Table 1. Metal films evaporated on to rocksalt. Brück (1936)

Metal	Plane $\parallel$ (100) NaCl	Orientation	% errors w.r.t. NaCl	Any better fit	Plane $\parallel$ (100) NaCl	Orientation	% errors w.r.t. NaCl	Remarks
Ni	(100)	Cube edges $\parallel$	$\pm 25.0\%$	Yes?	(210)	Ni edge $\parallel$ NaCl face diag.	-0.8% -11.2%	Only good fit in one direction. Symmetry relation lost
Cu	(100)	Cube edges $\parallel$	$\pm 28.0\%$	Yes?	(210)	Cu edge $\parallel$ NaCl face diag.	$+1.1\%$ -9.5%	Ditto.
Fe at two diff. temps.	(110)	Cube edges $\parallel$	-28.1% $\pm 1.4\%$	Yes	(100)	Cube edges $\parallel$	$+1.4\%$	Apparently a much better fit with no loss of symmetry relation
	(100)	NaCl edge $\parallel$ Fe face diag.	-28.1%	Yes				
Cr at two diff. temps.	(110)	Cube edges $\parallel$	-27.9% $+1.5\%$	Yes	(100)	Cube edges $\parallel$	$+1.5\%$	Ditto
	(100)	NaCl edge $\parallel$ Fe face diag.	-27.9%	Yes				
Al at two diff. temps.	(111)	A pair of face diags. $\parallel$	-24.0% -28.4%	Yes	(100)	Al edge $\parallel$ NaCl face diag.	$+1.2\%$	Ditto
	(100)	Cube edges $\parallel$	-28.4%	Yes				
Ag	(100)	Cube edges $\parallel$	-27.6%	Yes	(100)	Ditto.	$+2.4\%$	Ditto
Au at two diff. temps.	(111)	A pair of face diags. $\parallel$	$\pm 24.8\%$ -28.0%	Yes	(100)	Au edge $\parallel$ NaCl face diag.	$+2.1\%$	Ditto
	(100)	Cube edges $\parallel$	-28.0%	Yes				

Table 2. Metals deposited on beaten foils. Finch and Sun (1936)

Substrate	Deposit	Substrate orientation	Deposit plane surface	Other planes ;		% error w.r.t. substrate	Any better fit ?	Deposit plane surface	Other planes $\frac{h}{k}$		% error w.r.t. substrate	Remarks
				Sub.	Dep.				Sub.	Dep.		
Pt	Cu	(110)	(110)	($\bar{1}\bar{1}0$)	($\bar{1}\bar{1}0$)	- 7.6%	No	—	—	—	—	Symmetrical orientation
Pt	Ni	(110)	(110)	($\bar{1}\bar{1}0$)	($\bar{1}\bar{1}0$)	- 9.9%	No	—	—	—	—	Ditto
Pt	Co	(110)	(110)	($\bar{1}\bar{1}0$)	($\bar{1}\bar{1}0$)	- 9.1%	No	—	—	—	—	Ditto
Pd	Cu	(100)	(100)	($\bar{1}\bar{1}0$)	($\bar{1}\bar{1}0$)	- 6.7%	No	—	—	—	—	Ditto
Pd	Fe	(100)	(100)	(110)	(100)	+ 4.8%	No	—	—	—	—	Change of orientation due to body-centred deposit.
Au	Fe	(100)	(100)	(011)	(001)	- 0.6%	No	—	—	—	—	Ditto
Au	Co	(100)	(100)	(001)	(001)	- 12.8%	No	—	—	—	—	Symmetrical orientation
Au	Ni	(100)	(100)	(011)	(001)	- 22.3%	Yes	(100)	(100)	(100)	- 13.5%	Unusual since only face-centred metal showing this orientation

frequent twinning of the nickel when the deposit grows thicker. Chromium on copper showed two orientations in both of which important zone axes are parallel in the two crystals.

An example of the influence of faults on crystal growth is shown in a paper by Willems (1943) which studies the growth of urotropine on gypsum. The small parallelograms of the urotropine appear lined up with their sides parallel to steps in the gypsum.

While mere agreement in spacing is not conclusive, even when accompanied, as it must be, by similarity in arrangement of the atoms, there is no doubt from Royer's work that it plays an important part. If we take a figure of 12% as giving the maximum permissible disagreement for halides of mica, we can estimate the size of the embryo crystal which is sufficient to ensure indefinite growth under the conditions of the experiment. If the spacings of halide and mica are 1 and $1 + \alpha$, then after n steps the atoms will be αn apart. We may suppose that while this is

Table 3. Metallic deposits on copper. Cochrane (1936)

Metal	Orientation	% error w.r.t. Cu	Any better fit ?	Orientation	% error w.r.t. Cu	Remarks
Ni	Completely	— 2.4%	No	—	—	Takes Cu lattice constants for thin layer
Cu	Ditto	Nil	No	—	—	—
Zn	Does not orientate	—	—	—	—	—
Cd	Ditto	—	—	—	—	—
Ag	Completely	+13.2%	Yes	(111)'s . Rotated 30° w.r.t. one another.	—2.0%	Not complete symmetry
Cr	(110) Cr (111) Cu. Cube edge of Cr face diag. of Cu	+12.7% — 8.1%	No	—	—	Two possibilities of orientation
Co	Completely	— 1.6%	No	—	—	—

less than $\frac{1}{2}$, the forces on the whole tend to favour growth compared with a random position, but when it exceeds this value then they are unfavourable. Hence $\alpha n = \frac{1}{2}$, which for $\alpha = 0.12$ gives $n = 4$. The spacing in question being 5.15 Å, we get 20 Å. as the size of the embryo.

One might expect the necessary size of the embryo would decrease with increasing supersaturation, and it would be interesting to try if any effect of this kind can be observed.

Another kind of orientated growth is that which sometimes occurs when a crystal takes part in a chemical reaction. Many years ago I was able to show (Thomson 1930) that a single crystal of copper could oxidize to Cu_2O in such a way that the axes of the two crystals, both of which are cubic, were parallel. Since then a number of other examples have been found. In studying these we must remember that only certain faces of the original crystal will be exposed, and that, except in the case of cleavage faces, these are not necessarily the apparent faces,

since, for example, etching may produce pits covering the whole or almost the whole of the apparent surface. It is only on exposed faces that orientated growth can occur.

An examination of the observed instances allows certain general conclusions to be drawn. In general we have to do with a reaction in which new atoms are introduced, as when metals are oxidized, or in which some new atoms are substituted for old ones, as in the growth of one halide of silver on another, or silver bromide on thallous bromide, recently studied by Schwab (1947).

In most cases the atoms that remain keep their relative positions, though the distances between them necessarily often change. In some cases, e.g. growth of cuprous bromide and iodide on the (111) face of copper (Usmani 1941), the arrangement of atoms is similar in the two crystals but of different orientation. Here the copper atoms in both the metal and the halide form a lattice of equilateral triangles, but the two are turned with respect to one another through a right angle. The two planes appear always to have the same symmetry and general arrangement of atoms.

There is a preference for a fit in actual spacing where this is possible. Thus the rotation of the planes of the copper halides just mentioned greatly improves the fit, which is about 3% for the iodide and still better for the bromide, though it necessarily implies that not every atom has a partner in the new crystal.

The requirements of fit are not too exacting. Thus Schwab (1947) found hexagonal AgI can grow on the (111) face of AgBr with a difference of 13%, and some of the oxides show even larger differences, 18% for copper and 28% for barium (Burgers and Van Amstel 1936). Schwab has pointed out that the growth of AgBr on TlBr is probably helped by the fit normal to the plane of contact, which is (110) in both crystals, the large bromine atoms filling spaces left by the smaller metal atoms in the layer. The discrepancy in dimensions in this case is 2.5%.

There is much yet to be found out in the study of orientated growth. It is probable that different considerations are dominant in different cases. Equality of spacing is important but is not the only thing that matters. Crystal growth, and especially the part played in it by the Volmer layer, is a fascinating subject, which may be expected to increase greatly our knowledge of the other surface phenomena. The study of orientated crystals produced by chemical action is one of the most promising ways of finding out more about the mechanism of chemical change on which, after centuries of chemical study, science is still grossly ignorant.

REFERENCES

- AMSLER, J., 1942, *Helv. Phys. Acta*, **15**, 699.
ANDRADE, E. N. DA C., and MARTINDALE, J. G., 1935, *Phil. Trans. Roy. Soc. A*, **235**, 69.
BERG, W. F., 1938, *Proc. Roy. Soc. A*, **164**, 79.
BRANDES, H., 1927, *Z. phys. Chem.*, **126**, 196.
BRÜCK, L., 1936, *Ann. Phys., Lpz.*, **26**, 233.
BRÜCK, L., and LASSEN, H., 1935, *Ann. Phys., Lpz.*, **22**, 65.
BURGERS, W. G., and VAN AMSTEL, J. J. A. P., 1936, *Physica*, **3**, 1037.
COCHRANE, W., 1936, *Proc. Phys. Soc.*, **48**, 723.
COCKCROFT, J. D., 1928, *Proc. Roy. Soc. A*, **119**, 293.
FINCH, G. I., and SUN, C. H., 1936, *Trans. Faraday Soc.*, **32**, 852.
GIBBS, J. W., 1906, *Scientific Papers*, Vol. I, 324.
KOSSEL, W., 1928, *Phys. Z.*, **29**, 553; 1934, *Ann. Phys., Lpz.*, **21**, 457.
KOWARSKI, L., 1935, *J. Phys. Chem.*, **32**, 303, 395, 469.

- LASSEN, H., 1934, *Phys. Z.*, **35**, 172.
 LOTMER, W., 1945, *Helv. Phys. Acta*, **18**, 369.
 MARCELLIN, R., 1918, *Ann. Phys. Chem.*, **10**, 185.
 ROYER, L., 1928, *Bull. Soc. franç. Mineral.*, **51**, 7-159.
 RÜDIGER, O., 1937, *Ann. Phys., Lpz.*, **30**, 505.
 SCHWAB, G. M., 1947, *Trans. Faraday Soc.*, **128**, 641.
 STRANSKI, I. N., 1928, *Z. phys. Chem.*, **136**, 259; 1943, *Z. Kristallogr.*, **105**, 90.
 STRANSKI, I. N. and KAISCHEW, R., 1934, *Z. phys. Chem.*, **26**, Abt. B.1, 100, 312, Abt. B.4-5, 317; 1935, *Phys. Z.*, **36**, 393, *Ann. Phys. Lpz.*, **23**, 330; 1937, *Z. phys. Chem. B*, **35**, 427.
 THOMSON, G. P., 1930, *Proc. Roy. Soc. A*, **128**, 641.
 THOMSON, G. P., and COCHRANE, W., 1938, *Theory and Practice of Electron Diffraction* (London: Macmillan and Co.).
 USMANI, I. H., 1941, *Phil. Mag.*, **7**, 32, 89.
 VALETON, J. J. P., 1923, *Z. Kristallogr.*, **59**, 168, 335; 1924, *Ibid.*, **60**, 1.
 VOLMER, M., 1921, *Z. Elektrochem.*, **35**, 555.
 VOLMER, M., and ADHIRKARI, G., 1925, *Z. Phys.*, **35**, 170; 1926, *Z. phys. Chem.*, **119**, 46.
 VOLMER, M., and ESTERMANN, I., 1921, *Z. Phys.*, **7**, 13.
 VOLMER, M., and WEBER, A., 1926, *Z. phys. Chem.*, **119**, 277.
 WEST, C. D., 1945, *J. Opt. Soc. Amer.*, **35**, 26.
 WILLEMS, J., 1943, *Z. Kristallogr.*, **105**, 149.
 WULFF, G., 1901, *Z. Kristallogr.*, **34**, 449.

The Interpretation and Application of Electron-Diffraction "Kikuchi-Line" Patterns—

Part II. The Methods of Indexing the Patterns

By H. WILMAN

Applied Physical Chemistry Laboratories, Imperial College, London

MS. received 17 October 1947

ABSTRACT. Seven methods are described of indexing electron-diffraction Kikuchi-line patterns from single crystals for the purpose of practical applications of such patterns. The crystal lattice form and orientation can be determined by method 1 (see Part I), which also gives the indices of the diffractions relative to the chosen unit cell. The other six methods are in general only applicable in conjunction with a knowledge of the crystal lattice form and orientation, in which case any one of them, or more conveniently several in conjunction, can be used to index the lines.

§ 1. INTRODUCTION

THE only generally applicable direct method, independent of information about the crystal other than that contained in the diffraction pattern and a knowledge of the diffraction conditions, is the method described by Wilman (1948)—to be referred to as Part I—involving a construction of the reciprocal lattice. When a suitable unit cell in this reciprocal lattice has been chosen, the crystal lattice orientation and Laue symmetry can be defined, and the Laue indices of the Kikuchi-line diffractions follow simply as the coordinates of the corresponding reciprocal-lattice points, relative to the chosen reciprocal axes.

The other methods described here may be useful, for example, when too small a range of diffracted beams is recorded, or is distinguishable, to apply the absolute method described in Part I, or when the crystal nature and structure are already known but it is required to explore the intensities of a large range of diffractions, or

to define the orientation more precisely than can be done from the spot pattern when the spots are too diffuse or too few to apply the methods previously developed by the author (Charlesby, Finch and Wilman 1939).

One or other of these methods, or several in conjunction, can generally be applied successfully, so that the desired information about the relative orientations and relative reflecting powers of the different net planes can be derived and specified.

§ 2. OUTLINE OF THE MAIN FEATURES OF THE METHODS, AND SOME ESSENTIAL COMMON CONSIDERATIONS

(i) *The general characteristics of the methods*

The seven methods listed below are of three types. The first two methods are characterized by construction of a projection in which each net plane is represented by a point. It is only the first method, the reciprocal-lattice projection, which incorporates information on the net-plane spacings as well as on the directions of the net-plane normals represented in the second, gnomonic projection, method. The main feature of the second type of method is the derivation of the ratios of the net-plane or the Laue indices from the ratios of the intercepts cut off by the net-plane trace, i.e., the line-pair median, either on the three lines which are along the projections of the crystal axes on the plate (method 3), or (methods 4 and 5) on two perpendicular axes on the plate, one of which is along a definite ($h_1k_1l_1$) net-plane trace on the plate. The third type of method uses either the zone relationships (method 6) or the closely related parabolic and circular envelope relationships (method 7). All the methods are developed for the general case of a triclinic crystal, and the expressions applicable to crystals of other systems can be derived by giving $a, b, c, \alpha, \beta, \gamma$ appropriate values.

Methods 2, 4, 5 and 6 supply only the Miller indices of the reflecting net plane as a first step, and the order number of the Bragg reflection must then be found from the calculated plane spacing d and the measured line-pair separation D . Methods 1, 3 and 7, however, provide the Laue indices of the line diffractions directly, the measurements of D being used in combination with the other data which define the net-plane direction.

The first method alone actually carries out a construction of the reciprocal lattice and thus an independent determination of the crystal lattice, assuming only a knowledge of the camera length L and the position of O_1 , the foot of the perpendicular from the crystal origin O to the photographic plate. The wavelength must also be known if an estimation of the axial lengths a, b, c is required instead of just the axial ratios. Method 2 can also be used to a limited extent for a determination of the lattice form and orientation, but only in rather special cases where the crystal setting is a favourable one.

In all the other methods the further knowledge is assumed of the lattice constants $a, b, c, \alpha, \beta, \gamma$ and of the crystal orientation relative to the photographic plate or the primary beam. This information, if not definitely known or obtained by the method of Part I, can sometimes be inferred from the relative positions of the main bands in the pattern, possibly together with information provided by the geometrical form of the spot pattern and, if necessary, in conjunction with an examination of other patterns obtained at different settings, or patterns of other types such as powder and rotation patterns.

The various methods are mostly best suited to different spheres of application. Method 1 is the most fundamental, generally applicable and independent.

Method 2 is not usually so convenient as the other methods and is only readily usable when the polar direction of the projection is along or close to a crystal axis or a principal lattice row. Method 4 (Finch, Quarrell and Wilman 1935) was developed by the author to study the relationships between the Kikuchi lines and their parabolic and circular envelopes in reflection patterns at different azimuths; and method 5, also evolved in this laboratory (Thirsk and Whitmore 1940) is a form of method 4 more directly expressed for application to particular main azimuths. The neater and simpler method 3 was developed by the author in 1939 for easier practical use, especially when the crystal orientation is not very near to a main setting but is known relatively closely from the Laue spot positions as in the case of one of the anthracene patterns reproduced by Charlesby, Finch and Wilman (1939), where this method was first applied. Methods 6 and 7 are easy to apply and are often exceedingly useful, especially if only the intensified line of a pair is distinguishable; and when desirable these methods can be systematically applied in the way shown below.

(ii) *Preliminary considerations applicable to all methods*

The following considerations should be noted in indexing a pattern, whatever the method used.

(a) *Conventional viewpoint to be used in indexing prints.* The indexing of the pattern can be carried out on either a negative or a positive print; the latter is preferable for clearness. A positive print should be considered as corresponding to the electron distribution in the recording plane seen from the side of this plane opposite to that on which the electrons fall and on which the crystal lies (Part I). A negative print (or the initial "negative" recording observed from the sensitive emulsion side) should be regarded as corresponding to the electron distribution in the recording plane, viewed from the side on which the electrons fall and on which the crystal lies.

(b) *The Bragg-reflection positions at the band edges.* These have been shown in Part I to be at, or practically at, the points where the intensity is equal to the background intensity in the region considered.

(c) *The black and white line positions.* The positions of the Bragg reflection loci and the intersections of the net planes with the plate are determined by the orientation of the crystal relative to the plate, and are best specified relative to the foot O_1 of the normal from the crystal origin O to the plate. The primary electron beam OF is not in general along OO_1 , and of a Kikuchi-line pair the one farthest from OF is the intensified line and the one nearest OF the absorption line relative to the general background intensity.

(d) *Convention for the sign of the Laue indices of the line diffractions.* The direction of a net plane is specified equally well by the Miller indices (hkl) or $(-h, -k, -l)$, but in a Laue spot pattern the diffraction directions represented by the Laue indices hkl and $-h, -k, -l$ are different, and the sign of each index is fixed by the directions of this diffraction and of the primary beam relative to the corresponding crystallographic axis. In Kikuchi-line patterns, owing to the partial reciprocal relationships between the diffraction lines of a pair, there remains the question of which line of the pair is to be called the hkl diffraction and which the $-h, -k, -l$. It is important to adopt a convention to specify the sign to be used

for each line, not merely for uniformity, but in order to be able to use the zone and envelope Laue-index relationships (methods 6 and 7) without errors arising from this cause.

A convention which is easy to apply, and which is now recommended, is the following : Consider the two lines of a pair as the intersection with the plate of the two halves of a cone whose apex is at the crystal origin O , each line of the pair corresponding to one half of the cone on one side only of the net plane (also through O) whose trace on the plate is the median of the line pair. Then, having chosen the positive directions of the axes a, b, c , the positive (hkl) plane normal is that from the origin O to the plane which cuts off intercepts $a/h, b/k, c/l$ on a, b, c respectively, and the Kikuchi line which corresponds to the half-cone lying on the same side of O as this plane can be assigned the Laue indices hkl or the appropriate multiples nh, nk, nl . Conversely, the other line of the pair has Laue indices $-nh, -nk, -nl$ and corresponds to the half-cone lying on the opposite side of O , where also is the plane whose intercepts on a, b, c are $-a/h, -b/k, -c/l$ respectively. A paper or card model comprising the ab, bc , and ca faces of the unit cell is helpful in visualizing orientations of a particular net plane when the crystal orientation relative to the photographic plate is known.

The above convention is automatically followed if the indexing of the lines is done by method 1, but it has to be specifically applied in all the other methods of indexing.

(e) *Conversion of plane-trace and Kikuchi-line positions to those which would be obtained on a different recording plane.* The expressions derived below in methods 3, 4, 5 for the positions of the lines and the line-pair medians or plane normals take more convenient simple forms in the case where the known direction which defines the normal to the photographic plate has relatively small integral indices. If the plate normal OO_1 in any particular case does not approximately coincide, as it is usually arranged to do, with such a lattice row, the position that a line or a band median would have on a plate normal to any desired lattice row $[uvw]$ inclined at δ to the actual plate normal can be constructed as follows (figure 1).

Let O_2 be the (observed) intersection of the lattice row $[uvw]$ with the plate and O_2M be normal to O_1O_2 . Let QM be a line-pair median, i.e., a net-plane trace on the plate, produced to meet O_2M at M and O_2O_1 at Q . Then, as

can be seen from the side elevation in figure 1, the line OQ which lies in the net-plane under consideration is prolonged to meet at Q' the plane which is normal to OO_2 , i.e., $[uvw]$. If this new recording plane is now rotated about the line O_2M so as to become coincident with the actual plate, O_2Q' rotates to the position O_2Q'' ; whence the position of Q'' is as shown in the plan part of figure 1. In the triangle OO_2Q , let angle $O_1OO_2 = \delta$, angle $O_2OQ = \xi$, and $O_2Q = y$, then

$$y/\sin \xi = L'/\sin \{180^\circ - \xi - (90^\circ - \delta)\},$$

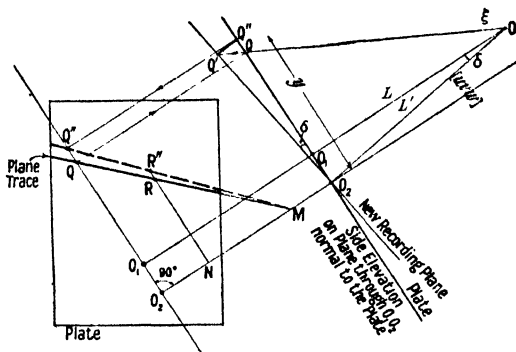


Figure 1.

and Q'' is given by

$$Q''O_2/QO_2 = (y + \Delta y)/y = (L'/y) \tan \xi = \cos \delta / \{1 - (y/L') \sin \delta\}.$$

The plane trace on the new recording plane is then the line $Q'M$, which becomes $Q''M$ when rotated into the plane of the plate. It follows that any other ordinate RN is also changed to $R''N$ where $R''N/RN = Q''O_2/QO_2$.

As a particular example it can be seen that if δ is less than about 5° and $L' \sim 20$ cm. with $RN < 6$ cm., then $\Delta y/y < 2\%$, thus under these conditions, to a first approximation, the pattern can be indexed as if OO_2 was the normal to the plate.

The new Kikuchi-line positions can be found, if necessary, by treating as plane appropriate small parts of the corresponding diffraction cones, and applying the above construction.

§ 3. THE METHODS OF INDEXING THE PATTERNS

(i) *The reciprocal-lattice projection method*

This has been fully described and illustrated in Part I and will not be discussed further here.

(ii) *The gnomonic projection method*

A gnomonic projection can be constructed very easily from the positions of the band medians. If any net plane is shown in side view by the line OP in figure 2

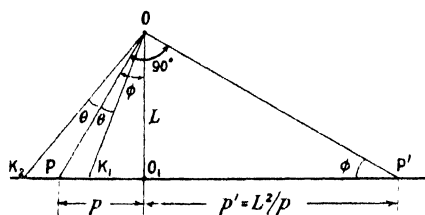


Figure 2. Gnomonic projection.

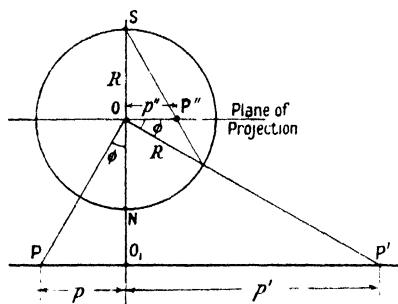


Figure 3. Stereographic projection.

(K_1 and K_2 representing a Kikuchi-line pair from this plane), then the plane normal is OP' , $O_1P' = L \cot \phi = L^2/p$; ϕ is the inclination of the net plane to OO_1 . The gnomonic projection on the recording plane comprises points such as P' where the plane normals meet this plane, and it is thus simply obtained by marking the position of O_1 on the pattern and then marking a point corresponding to each band median, at a distance proportional to $1/p$ from O_1 on a line through O_1 normal to the band median, and on the opposite side of O_1 from the band median. The application to x-ray diffraction patterns has been illustrated by Wyckoff (1931), Bragg (1933), and others.

Although stereographic projections are not easy to apply to the indexing of Kikuchi-line patterns, it may sometimes be desired to construct such a projection, and this can be done as follows. Representing a net plane in edge view by OP (figure 3) and its normal by OP' , meeting the plate at P' , its stereographic projection on a plane through O parallel to the plate is P'' , of radius R . By joining the point where OP' cuts the sphere to the opposite pole S , the point P'' is found as the

intersection of this line with the equatorial projection plane. It is easy to show that the length OP'' , i.e. p'' , is

$$p'' = R \cos \phi / (1 + \sin \phi) = R \{ [1 + (p/L)^2]^{\frac{1}{2}} + p/L \}^{-1}.$$

To construct a stereographic projection from a Kikuchi line pattern it is thus necessary to measure off from O_1 (or a point displaced suitably outside the pattern area) distances p'' in a direction opposite to that of the normal drawn from O_1 to each line-pair median.

(iii) *The method of intercepts on the three axial projections on the plate*

The Laue indices of the Kikuchi lines can be calculated directly from the bandwidth D and the intercepts which the band median cuts off on the projections $\mathbf{a}_p, \mathbf{b}_p, \mathbf{c}_p$ of the crystal axes $\mathbf{a}, \mathbf{b}, \mathbf{c}$, drawn from the point O_1 where a known lattice row $[uvw]$ meets the plate normally (or nearly so). In figure 4, let O be the origin in the crystal and OO_1 , of length L , be the normal from O to the plate. Let $\mathbf{a}, \mathbf{b}, \mathbf{c}$ meet the plate at A, B, C , respectively, so that O_1A, O_1B, O_1C are in the directions of $\mathbf{a}_p, \mathbf{b}_p, \mathbf{c}_p$ and let the (hkl) plane through O cut O_1A, O_1B, O_1C at X, Y, Z , distant x, y, z from O_1 . Also let $\mathbf{N}$ be the unit vector along OO_1 and $\mathbf{n}$ that along the normal from O outwards to the plane whose intercepts on $\mathbf{a}, \mathbf{b}, \mathbf{c}$ are $\mathbf{a}/h, \mathbf{b}/k, \mathbf{c}/l$; the angles between $\mathbf{N}$ and $\mathbf{a}, \mathbf{b}, \mathbf{c}$ are $\theta_a, \theta_b, \theta_c$ respectively. Then we have

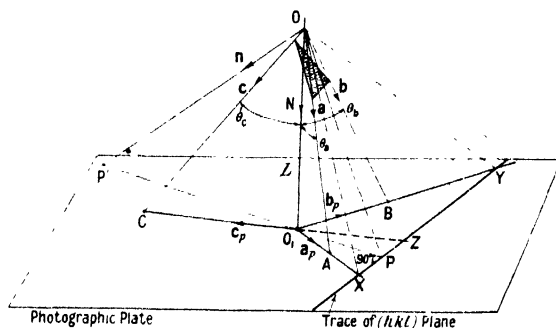


Figure 4.

$\mathbf{O}_1\mathbf{A} = \mathbf{OA} - \mathbf{OO}_1 = (L/\cos \theta_a)(\mathbf{a}/a) - L\mathbf{N}$. The equation of the line O_1A relative to O as origin is thus

$$\mathbf{r} = L\mathbf{N} + mL\{(\mathbf{a}/a \cos \theta_a) - \mathbf{N}\},$$

where m is a parameter. Since the point X on this line lies in the (hkl) plane through O , its radius vector $\mathbf{r}$ must also satisfy $\mathbf{r} \cdot \mathbf{n} = 0$.

Hence, with $\mathbf{a} \cdot \mathbf{n} = hd$,

$$L\mathbf{N} \cdot \mathbf{n} + mL\{[(\mathbf{a} \cdot \mathbf{n})/a \cos \theta_a] - \mathbf{N} \cdot \mathbf{n}\} = 0,$$

i.e.

$$m = \cos \theta_p / \{\cos \theta_p - (hd/a \cos \theta_a)\},$$

where θ_p is the angle between $\mathbf{N}$ and $\mathbf{n}$. We now have, directly, $O_1A = L \tan \theta_a$,

whence $O_1X = x + mO_1A = L \tan \theta_a \cos \theta_p \{ \cos \theta_p - (hd/a \cos \theta_a) \}$

and

$$h = (a/d) \cos \theta_p \cdot [\cos \theta_a - (L/x) \sin \theta_a],$$

and there are similar expressions for k and l .

The values of d and $\cos \theta_p$ must be expressed in terms of the bandwidth D , the Laue indices hkl of the diffraction line, and the length p of the perpendicular from O_1 to the trace XYZ of the (hkl) plane (subtending $90^\circ - \theta_p$ at O), thus

$$d = \lambda / (2 \sin \theta) \simeq \lambda L / D \sin \theta_p; \quad \cos \theta_p = p / (L^2 + p^2)^{\frac{1}{2}}; \quad \sin \theta_p = L / (L^2 + p^2)^{\frac{1}{2}}.$$

Hence $h = [pD/(L^2 + p^2)](a/\lambda)[\cos \theta_a - (L/x) \sin \theta_a]$.

Similarly $k = [pD/(L^2 + p^2)](b/\lambda)[\cos \theta_b - (L/y) \sin \theta_b]$,

$$l = [pD/(L^2 + p^2)](c/\lambda)[\cos \theta_c - (L/z) \sin \theta_c].$$

In order to calculate h , k and l for a line pair it is thus necessary to calculate θ_a , θ_b and θ_c and the directions of $\mathbf{a}_p$, $\mathbf{b}_p$, $\mathbf{c}_p$ for the known or estimated crystal setting (actually in most main settings these are almost obvious or very simply calculated), and to measure λL , x , y , z , p and D and then use these equations. λ need not be known if only $h : k : l$ is accepted as sufficient to determine h , k and l . The lengths x , y , z are positive or negative according to whether they are measured from O_1 along the "positive" or the "negative" direction of $\mathbf{a}_p$, $\mathbf{b}_p$ or $\mathbf{c}_p$. The signs of h , k and l also depend on p and θ_p . If p is always positive and thus θ_p always an acute angle, i.e. to the positive direction of the net-plane normal n always down from O towards the plate, then with the index convention stated in § 2 the indices h , k , l are assigned to the line diffraction (black or white) which lies on the side of the plane trace XYZ nearest to O_1 , and the line on the opposite side XYZ has the indices with all three signs reversed.

When only one line of the pair is visible, so that p and D cannot be measured, the bandwidth D can be estimated roughly from the sharpness of the line, or much more exactly from its position in relation to other lines which form with the line concerned a parabolic envelope with the associated zone axis intersection at its focus, so that $D/2$ can be measured.

If $p = 0$, and thus $x = y = z = 0$, the only finite terms remaining in the equations for h , k and l are those containing the factors p/x , p/y and p/z which are the cosines of the angles between O_1P and O_1A , O_1B and O_1C respectively.

Calculation of θ_a , θ_b , θ_c when OO_1 is along $[uvw]$. If the angles θ_a , θ_b , θ_c are not directly known, but a lattice row $[uvw]$ is known to be along OO_1 , the values of $\cos \theta_a$, $\sin \theta_a$ etc. can be calculated as follows for insertion in the above equations for h , k and l .

If $\mathbf{T}$ is the (positive) $[uvw]$ lattice translation,

$$\cos \theta_a = \mathbf{T} \cdot \mathbf{a} / Ta = (ua + vb \cos \gamma + wc \cos \beta) / T,$$

$$\cos \theta_b = \mathbf{T} \cdot \mathbf{b} / Tb = (ua \cos \gamma + vb + wc \cos \alpha) / T,$$

$$\cos \theta_c = \mathbf{T} \cdot \mathbf{c} / Tc = (ua \cos \beta + vb \cos \alpha + wc) / T,$$

where $T^2 = u^2a^2 + v^2b^2 + w^2c^2 + 2uvab \cos \gamma + 2vwb \cos \alpha + 2wuc \cos \beta$.

Since $\sin^2 \theta_a = 1 - \cos^2 \theta_a$, this gives

$$\sin^2 \theta_a = [v^2b^2 \sin^2 \gamma + w^2c^2 \sin^2 \beta + 2vwb \cos \alpha - \cos \beta \cos \gamma] / T^2,$$

$$\sin^2 \theta_b = [u^2a^2 \sin^2 \gamma + w^2c^2 \sin^2 \alpha + 2uwac \cos \beta - \cos \gamma \cos \alpha] / T^2,$$

$$\sin^2 \theta_c = [u^2a^2 \sin^2 \beta + v^2b^2 \sin^2 \alpha + 2uvab \cos \gamma - \cos \alpha \cos \beta] / T^2.$$

Thus, for example,

$$h = \{pD/(L^2 + p^2)\}(a/\lambda)(1/T) \cdot [(ua + vb \cos \gamma + wc \cos \beta) - (L/x)\{v^2b^2 \sin^2 \gamma + w^2c^2 \sin^2 \beta + 2vwb \cos \alpha - \cos \beta \cos \gamma\}^{1/2}].$$

These expressions take much simpler forms in the more symmetrical crystal systems, thus for example for cubic crystals,

$$h = [pD/(L^2 + p^2)](a/\lambda)(u^2 + v^2 + w^2)^{-1/2}[u - (L/x)(v^2 + w^2)^{1/2}],$$

$$k = [pD/(L^2 + p^2)](b/\lambda)(u^2 + v^2 + w^2)^{-1/2}[v - (L/y)(u^2 + w^2)^{1/2}],$$

$$l = [pD/(L^2 + p^2)](c/\lambda)(u^2 + v^2 + w^2)^{-1/2}[w - (L/z)(u^2 + v^2)^{1/2}].$$

Calculation of the angles between $\mathbf{a}_p$, $\mathbf{b}_p$, and $\mathbf{c}_p$ and the normal to the trace of the $(h_1k_1l_1)$ plane passing through $[uvw]$, normal to the plate. When the directions of the projections $\mathbf{a}_p$, $\mathbf{b}_p$, $\mathbf{c}_p$ are not conveniently determined by inspection from the known orientation or the geometry of the spot pattern, i.e. the Laue zones along the spot rows, they can be calculated from the known indices $[uvw]$ of OO_1 and those $(h_1k_1l_1)$ of a plane which passes through this zone axis and whose Kikuchi lines are identified in the pattern.

Thus, the vector $\mathbf{r}_1^* = h_1\mathbf{a}^* + k_1\mathbf{b}^* + l_1\mathbf{c}^*$ lies in the plate, normal to $(h_1k_1l_1)$, and since $\mathbf{a}_p = \mathbf{a} - (\mathbf{a} \cdot \mathbf{N})\mathbf{N}$,

$$\cos(\mathbf{a}_p, \mathbf{r}_1^*) = (\mathbf{a}_p \cdot \mathbf{r}_1^*) / a_p r_1^* = (d_1/a_p) \{ \mathbf{a} \cdot \mathbf{r}_1^* - (\mathbf{a} \cdot \mathbf{N})(\mathbf{N} \cdot \mathbf{r}_1^*) \} = h_1 d_1 / a \sin \theta_a,$$

$$\cos(\mathbf{b}_p, \mathbf{r}_1^*) = k_1 d_1 / b \sin \theta_b, \quad \text{and} \quad \cos(\mathbf{c}_p, \mathbf{r}_1^*) = l_1 d_1 / c \sin \theta_c,$$

where $\sin \theta_a$, $\sin \theta_b$ and $\sin \theta_c$ are given by the expression stated above, in terms of u , v and w . For cubic crystals, for example, $\cos(\mathbf{a}_p, \mathbf{r}_1^*) = h_1 / (h_1^2 + k_1^2 + l_1^2)^{1/2} \sin \theta_a$, etc.

It is desirable to check the relative positive directions of $\mathbf{a}_p$, $\mathbf{b}_p$, $\mathbf{c}_p$ carefully, preferably with the aid of a rough model, before proceeding to index the pattern.

(iv) *The method of intercepts on two perpendicular axes along and perpendicular to the trace of a plane $(h_1k_1l_1)$, in terms of the azimuth ϕ in this plane*

An analytical derivation of the equation of a net-plane trace or Kikuchi line has been given and illustrated previously (Finch, Quarrell and Wilman 1935), and it can also be obtained in a similar way by vector methods. The results are as follows.

First imagine the crystal unit cell to be mounted in a fixed, but at present unspecified, position relative to three orthogonal reference axes OX_3 , OY_3 , OZ_3 . (A convenient setting is that shown in figure 5.) Now transfer the cell, together

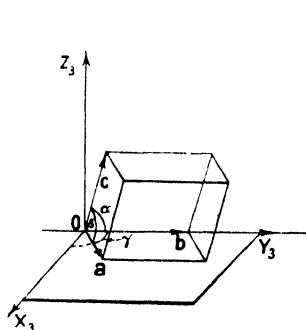


Figure 5.

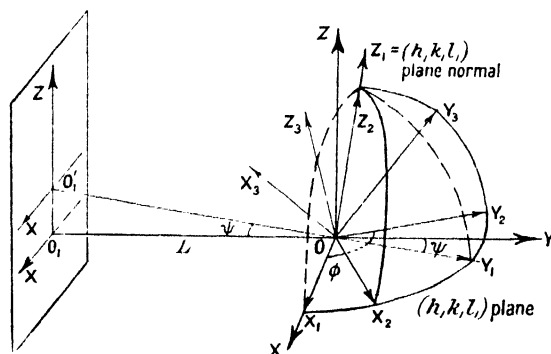


Figure 6.

with these axes, to the diffracting crystal, and orientate it so that the axes $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ of the unit cell are parallel to the corresponding axes of the crystal, whose origin is at O (figure 6). Let OO_1 be the normal to the fluorescent screen or photographic plate and assume the known crystal orientation to be defined by the indices $(h_1k_1l_1)$ of a net-plane whose trace on the plate is the line $O_1'X$, and by the azimuth angle ϕ between two planes normal to $(h_1k_1l_1)$, one also normal to the plate and thus containing OO_1 and OO_1' , and the other containing OY_3 and intersecting $(h_1k_1l_1)$ in OY_2 .

The orthogonal set of axes OX_2, OY_2, OZ_2 is further defined by setting OZ_2 along the positive normal to the net plane $(h_1k_1l_1)$, i.e. from O to the plane which cuts off intercepts $a/h_1, b/k_1, c/l_1$ on a, b, c respectively. A further orthogonal set of axes OX_1, OY_1, OZ_1 is such that OZ_1 coincides with OZ_2 , and OX_1 is parallel to the plate and lies in the $(h_1k_1l_1)$ plane through O , i.e. $OX_2Y_1Y_2$. Finally the orthogonal axes OX, OY, OZ are such that OX is along OX_1 and OY is along O_1O ; the angle YOY_1 is ψ .

On this basis, if the normals to the net planes $(h_1k_1l_1)$ and any other plane (hkl) have direction cosines λ_1, μ_1, ν_1 and λ, μ, ν relative to $OX_3Y_3Z_3$, the equation of the line where the plane (hkl) through O intersects the plate is, relative to O_1X and O_1Z on the plate (figure 6),

$$C_1z = -A_1x + B_1L,$$

where

$$A_1 = A_2, \quad A_2 = A_3 \cos \phi - B_3 \sin \phi,$$

$$B_1 = B_2 \cos \psi + C_2 \sin \psi, \quad B_2 = B_3 \sin \phi + B_3 \cos \phi,$$

$$C_1 = C_2 \cos \psi - B_2 \sin \psi, \quad C_2 = C_3,$$

$$A_3 = (\lambda\nu_1 - \nu\lambda_1)/(1 - \mu^2)^{\frac{1}{2}},$$

$$B_3 = \{\mu - \mu_1(\lambda\lambda_1 + \mu\mu_1 + \nu\nu_1)\}/(1 - \mu^2)^{\frac{1}{2}},$$

$$C_3 = \lambda\lambda_1 + \mu\mu_1 + \nu\nu_1,$$

in which

$$1 = \lambda^2 + \mu^2 + \nu^2 = \lambda_1^2 + \mu_1^2 + \nu_1^2 = A_1^2 + B_1^2 + C_1^2 = A_2^2 + B_2^2 + C_2^2 = A_3^2 + B_3^2 + C_3^2.$$

If the origin of coordinates on the plate is transferred from O_1 to O_1' , which is on the $(h_1k_1l_1)$ plane trace where it cuts O_1Z , the (hkl) plane trace has the equation $C_1z = -A_1x + B_2L/\cos \psi$, and if ψ is small this is approximately $C_2z = -A_2x + B_2L$, which is the same as the equation for the plane trace relative to O_1XZ with ψ put equal to zero.

The equation of the n th order Kikuchi-line pair from the (hkl) plane is, relative to O_1X and O_1Z on the plate, the hyperbola

$$(A_1x + B_1y + C_1z)^2 = (x^2 + y^2 + L^2)(n\lambda/2d)^2,$$

and when $x^2 + y^2$ is small compared with L^2 this approximates closely to the pair of parallel straight lines, $C_1z = -A_1x + \{B_1L - (n\lambda L/2d)\}$. The equation relative to O_1' is obtained analogously from the above equation for the plane trace, and when ψ is small it is effectively $C_2z = -A_2x + \{B_2L \pm (n\lambda L/2d)\}$.

If the crystal setting and lattice constants are known, these expressions can be evaluated, for example as shown below. Since the equation of the (hkl) plane trace involves h, k and l only linearly, the two intercepts of any line-pair median on O_1X and O_1Z (or one intercept and the slope) suffice to allow the ratios $h : k : l$ to be determined. The angle ϕ is taken here as the angle through which the crystal must be rotated in a positive direction (right-handed screw) about the positive direction of the $(h_1k_1l_1)$ plane normal, to bring the crystal from the setting where OY_2 is in the plane YOZ , into the actual crystal orientation.

The axes $OX_3Y_3Z_3$ can be chosen conveniently relative to a, b, c as shown in figure 4, such that OY_3 is along b and OX_3 is in the ab plane on the same side of b as is a . In this case, for the general triclinic crystal,

$$\lambda = [(h/a) + (k \cos \gamma)/b]d/\sin \gamma; \quad \mu = kd/b;$$

$$\nu = \{(h/a)(\cos \alpha \cos \gamma - \cos \beta) + (k/b)(\cos \beta \cos \gamma - \cos \alpha) + (l/c) \sin^2 \gamma\}abcd/V \sin \gamma,$$

where $d = (V/abc) / \{ \Sigma(h^2/a^2) \sin^2 \alpha + 2\Sigma(hk/ab)(\cos \alpha \cos \beta - \cos \gamma) \}^{\frac{1}{2}}$,

and $V = abc(1 - \cos^2 \alpha - \cos^2 \beta - \cos^2 \gamma + 2 \cos \alpha \cos \beta \cos \gamma)^{\frac{1}{2}}$.

There are similar expressions for λ_1, μ_1, ν_1 in terms of h_1, k_1, l_1 .

The angle ϕ is not in general known directly as part of the specification of the crystal orientation, but it can be calculated for example from the indices $[u_1 v_1 w_1]$ of the lattice row which is along OO_1 in the $(h_1 k_1 l_1)$ plane, since then we have

$$\cos \phi = (\cos \delta) / \mu_1 = (u_1 a \cos \gamma + v_1 b + w_1 c \cos \alpha) / \mu_1 T_1,$$

where δ is the angle between $[u_1 v_1 w_1]$ and OY_2 , i.e. b , and T_1 is the $[u_1 v_1 w_1]$ lattice translation given by $T_1^2 = \Sigma u_1^2 a^2 + 2 \Sigma u_1 v_1 ab \cos \gamma$.

Thus in any particular case, knowing the crystal lattice constants, the orientation defined by the indices $(h_1 k_1 l_1)$ and $[u_1 v_1 w_1]$ and the angle ψ which can be measured from the pattern in terms of the position of the $(h_1 k_1 l_1)$ line-pair median relative to O_1 , the procedure to index the pattern is as follows. First write down the values of λ_1, μ_1, ν_1 and λ, μ, ν from the above expressions by inserting the known indices h_1, k_1, l_1 , and $a, b, c, d, V, \alpha, \beta, \gamma$, and calculate $\cos \phi$ and $\sin \phi$, as just described, from $[u_1 v_1 w_1]$. Then putting these values in the expressions for A_2, B_2, C_2 etc., the equation for the (hkl) line median is obtained, involving h, k, l linearly. Finally the position of each line-pair median in the pattern can be compared with this equation and two relationships between h, k and l thereby established which give $h : k : l$; from this the simplest integral value of h, k and l consistent with the bandwidth D must be deduced, bearing in mind the accuracy of the measurements.

This method, like the next, is unwieldy for crystals which have low symmetry, but it has been easily applied to the patterns from some cubic and hexagonal crystals (Finch, Quarrell and Wilman 1935, Finch and Wilman 1936, 1937 a, b). For cubic crystals, for example, the expressions take the simple general forms

$$\lambda = hd/a, \quad \mu = kd/a, \quad \nu = ld/a, \quad d = a/(h^2 + k^2 + l^2)^{\frac{1}{2}},$$

$$\lambda_1 = h_1 d_1/a, \quad \mu_1 = k_1 d_1/a, \quad \nu_1 = l_1 d_1/a, \quad d_1 = a/(h_1^2 + k_1^2 + l_1^2)^{\frac{1}{2}},$$

$$A_3 = (hl_1 - lh_1)/(h^2 + k^2 + l^2)^{\frac{1}{2}}(h_1^2 + l_1^2),$$

$$B_3 = [k - k_1(hh_1 + kk_1 + ll_1)]/[(\Sigma h^2)(\Sigma h_1^2)(h_1^2 + l_1^2)^{\frac{1}{2}}],$$

$$C_3 = (hh_1 + kk_1 + ll_1)/[(\Sigma h^2)(\Sigma h_1^2)^{\frac{1}{2}}],$$

$$\cos \phi = [v_1/(\Sigma u_1^2)^{\frac{1}{2}}][\Sigma h_1^2/(h_1^2 + l_1^2)]^{\frac{1}{2}}.$$

The general analytical expressions derived above are also useful in studying the positional relationships between systems of net-plane traces or Kikuchi lines, such as the line envelopes.

(v) *The method of intercepts on two perpendicular axes along and perpendicular to the trace of an $(h_1 k_1 l_1)$ plane in which $[u_1 v_1 w_1]$ is normal to the plate*

Thirsk and Whitmore's (1940) expressions are given below, in a notation uniform with the above, and the equation of a line pair is added.

Figure 7 shows the disposition of the reference axes OX, OY and of the known plane $(h_1 k_1 l_1)$ normal to the plate, with positive direction of the known lattice row $[u_1 v_1 w_1]$ along the normal OO_1 from the crystal origin to the plate. The trace MN

Calculation of the position of the intersection of a zone-axis $[uvw]$ with the plate. Any lattice row $[uvw]$ along OP (figure 8), with lattice translation $\mathbf{T}$, makes with the plate normal OO_1 along the reciprocal-lattice direction $[h_0k_0l_0]^*$, i.e. $\mathbf{r}_0^*$, an angle η given by $\cos \eta = (\mathbf{T} \cdot \mathbf{r}_0^*) / T r_0^* = (uh + vk + wl)d_0 / T$. The radial distance of P from O_1 in the plate is then $O_1P = L \tan \eta$. If the direction O_1X of the trace

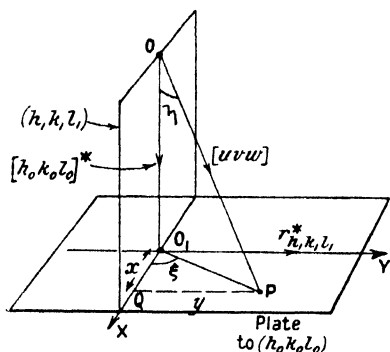


Figure 8.

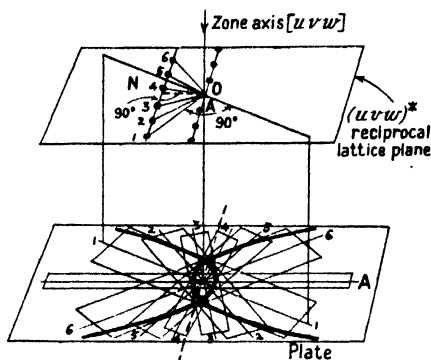


Figure 9.

of a known plane $(h_1k_1l_1)$ through OO_1 is identified, and PQ is the normal from P to O_1X , then O_1P makes with O_1X an angle ξ_1 given by

$$\sin \xi_1 = QP / O_1P = (\mathbf{OP} \cdot \mathbf{r}_1^*) / r_1^* O_1P = (uh_1 + vk_1 + wl_1)d_1 / T \sin \eta.$$

The distance O_1P and the angle ξ_1 thus specify the point P where $[uvw]$ meets the plate. In the above expressions, $d_1 \simeq \lambda L / D_1$, where D_1 is the measured $h_1k_1l_1$ line-pair separation near O_1 . From the above equations it follows also that $y_1 = QP = O_1P \sin \xi_1 = L \{ (\sum uh_1) / \sum uh_0 \} d_1 / d_0$.

Calculation of the indices $[uvw]$ of a zone axis from the position of its intersection P with the plate. From the above equations, conversely,

$$uh_0 + vk_0 + wl_0 = (T/d_0) \cos \eta; \quad uh_1 + vk_1 + wl_1 = (T/d_1) \sin \xi_1 \sin \eta;$$

but to determine $[uvw]$ uniquely it is necessary to measure also the angle ξ_2 made by O_1P with the trace of another net-plane, for example $(h_2k_2l_2)$, through O_1 , whence

$$uh_2 + vk_2 + wl_2 = (T/d_2) \sin \xi_2 \sin \eta.$$

Then $u : v : w =$

$$\left| \begin{array}{ccc} (\cos \eta)/d_0 & k_0 & l_0 \\ (\sin \xi_1 \sin \eta)/d_1 & k_1 & l_1 \\ (\sin \xi_2 \sin \eta)/d_2 & k_2 & l_2 \end{array} \right| : \left| \begin{array}{ccc} h_0 & (\cos \eta)/d_0 & l_0 \\ h_1 & (\sin \xi_1 \sin \eta)/d_1 & l_1 \\ h_2 & (\sin \xi_2 \sin \eta)/d_2 & l_2 \end{array} \right| : \left| \begin{array}{ccc} h_0 & k_0 & (\cos \eta)/d_0 \\ h_1 & k_1 & (\sin \xi_1 \sin \eta)/d_1 \\ h_2 & k_2 & (\sin \xi_2 \sin \eta)/d_2 \end{array} \right|.$$

Alternatively, from the above expression for y_1 (and y_2), we have the following two equations, if $\rho_1 = d_1/d_0$ and $\rho_2 = d_2/d_0$:

$$u\{h_0y_1 - h_1L\rho_1\} + v\{k_0y_1 - k_1L\rho_1\} + w\{l_0y_1 - l_1L\rho_1\} = 0,$$

$$u\{h_0y_2 - h_2L\rho_2\} + v\{k_0y_2 - k_2L\rho_2\} + w\{l_0y_2 - l_2L\rho_2\} = 0,$$

which gives $u : v : w$ as in the case of a pair of zone-relation equations given above.

Calculation of $[h_0k_0l_0]^*$ if this direction is initially specified by $[u_0v_0w_0]$. If OO_1 is initially known to be the $[u_0v_0w_0]$ lattice direction, then the indices $[h_0k_0l_0]^*$ of this direction in the reciprocal lattice, required in the above calculations, may be obtained readily by comparison of the two equations

$$\mathbf{T}_0^* = h_1\mathbf{a}^* + k_1\mathbf{b}^* + l_1\mathbf{c}^* ;$$

$$\mathbf{T}_0 = u_0\mathbf{a} + v_0\mathbf{b} + w_0\mathbf{c} = (\mathbf{T}_0 \cdot \mathbf{a})\mathbf{a}^* + (\mathbf{T}_0 \cdot \mathbf{b})\mathbf{b}^* + (\mathbf{T}_0 \cdot \mathbf{c})\mathbf{c}^*.$$

Thus,

$$h_0 = \mathbf{T}_0 \cdot \mathbf{a} = a(u_0 a + v_0 b \cos \gamma + w_0 c \cos \beta) ;$$

$$k_0 = \mathbf{T}_0 \cdot \mathbf{b} = b(u_0 a \cos \gamma + v_0 b + w_0 c \cos \alpha) ;$$

$$l_0 = \mathbf{T}_0 \cdot \mathbf{c} = c(u_0 a \cos \beta + v_0 b \cos \alpha + w_0 c).$$

(vii) *The envelope-relation method*

All the net planes such as (hkl) which are parallel to the same zone axis $[uvw]$ intersect the plate in lines which are along the corresponding Kikuchi-line pair medians, and they all pass through one point, where the line $[uvw]$ drawn through the crystal origin O meets the plate. The indices (hkl) of all planes of the zone satisfy the relation $uh + vk + wl = 0$, but the Kikuchi lines of the group of cozoneal planes fall into subgroups, the lines of each subgroup being tangential to a common parabolic envelope.

The envelope curves themselves are not in general present in the pattern as true parabolic curves (black or white), though in the main subgroups where a large number of lines touch the envelope locus at frequent intervals, the polygonal nature of the apparent curves is only recognizable if the crystal is highly perfect and if a narrow well-focused monochromatic electron beam is used to obtain high resolution (Finch, Quarrell and Wilman 1935, Shinohara 1935, Finch and Wilman 1937b). There is usually noticeable a characteristic weakening and curving away of Kikuchi lines just before they touch the envelope, also contributing to the prominence and appearance of continuity of the main envelope loci.

As Shinohara (1935) has shown, the lines associated with any one parabola arise from net planes such as (hkl) belonging to a zone $[uvw]$; the reciprocal-lattice points associated with these planes not only lie on the $(uvw)^*$ plane through the origin O , fulfilling the condition $uh + vk + wl = 0$, but also lie on a line, for example in an $[h_1k_1l_1]^*$ direction, lying in the $(uvw)^*$ plane parallel to the plate or nearly so, but not passing through the origin. (If it did, the lattice points along it would only correspond to the orders of diffraction from a single net-plane.) If this reciprocal-lattice line $[h_1k_1l_1]^*$ is densely populated, the approximation to the parabolic envelope curve will be prominent. It can be shown that, if the zone axis is normal to the plate, or nearly so, the axis of the parabola is parallel to the perpendicular ON to this reciprocal-lattice line $[h_1k_1l_1]^*$, and is thus, in the plate, along the trace of the net-plane $(h_1k_1l_1)$ (figure 9). The Laue indices of the successive Kikuchi lines touching the parabola differ by the same integers $h_1k_1l_1$ as the indices of this net-plane, or by a multiple of them.

Thus, if the indices of two or more lines touching a particular parabolic locus are known, those of the plane $(h_1k_1l_1)$ along its axis are the differences of these two sets or a sub-multiple of these differences; the indices of all the other lines touching the locus are also easily found even if only one line of each pair is distinguishable, and can be checked by the position or slope of the line (see methods 3, 4, 5).

Conversely, if the plane ($h_1k_1l_1$) whose trace is along the parabola axis has been indexed and the indices hkl of any line touching the parabolic locus are known, then those of the other lines are of the form $h + nh_1$, $k + nk_1$, $l + nl_1$.

If the sign convention of §2 is adopted in assigning the Laue indices hkl to any Kikuchi line, black or white, the above reciprocal-lattice interpretation shows (figure 9) that Kikuchi lines touching a parabolic locus on one side of its axial net-plane progressively further from its vertex have indices obtained from those of the previous one by *adding* the Laue indices (or a multiple) of the first-order Kikuchi line due to the axial plane of the parabola and lying on the same side of the net-plane trace, i.e. the parabola axis.

As has been shown (Finch, Quarrell and Wilman 1935), the parabolic envelopes in turn fall into groups whose envelopes are a system of circles with centres at the points where main lattice rows, if produced, would meet the plate. Shinohara (1935) has also shown that the circular envelopes represent the limiting loci outside which lie all Kikuchi lines characterized by the same Laue index h' relative to a lattice row meeting the plate at the centre of the circle, the lines all lying on the side of the net-plane traces, opposite to that on which the centre of the circle lies. The angular radii of such circles are given by $\cos \phi = 1 - (h'\lambda/T)$, i.e. $\phi \simeq (2h'\lambda/T)^{\frac{1}{2}}$, where T is the lattice translation of the lattice row directed towards their centre and h' is any integer; thus the circles have radii nearly proportional to $\sqrt{h'}$, and have the same positions as the h' -th-order Laue zones of reinforcement from the lattice row corresponding to an imaginary primary beam parallel to the row.

If the Laue indices of a Kikuchi line are hkl relative to the axes **a**, **b**, **c**, and the lattice row is $[uvw]$, the line belongs to the group associated with the h' -th-order circle round the $[uvw]$ trace on the plate, where h' is the Laue index of the diffraction direction hkl relative to the lattice row $[uvw]$, that is $hu + kv + lw = h'$. Many of the lines of a group will not touch the circle nor even lie close to it, but they fall into subgroups as above, touching parabolic envelopes which do touch the circle; the connection between any particular line and one of these circles is thereby established, and if $[uvw]$ and h' are known, the above linear relation between h , k and l is obtained. A line may be seen to belong to two or more non-coaxial parabola or circles, and in this case several relations such as the above hold simultaneously and allow hkl to be determined without difficulty, on this basis alone, without reference to the bandwidth.

ACKNOWLEDGMENTS

The author thanks the Council of the Royal Society for a Moseley Research Studentship held from 1938–1940, during which period the third of the above methods was developed.

It is a pleasure also to thank Professor G. I. Finch for his interest and encouragement.

REFERENCES

- BRAGG, W. H., and BRAGG, W. L., 1933, *The Crystalline State*, Vol. I (London: G. Bell and Sons).
 CHARLESBY, A., FINCH, G. I., and WILMAN, H., 1939, *Proc. Phys. Soc.*, **51**, 479.
 FINCH, G. I., QUARRELL, A. G., and WILMAN, H., 1935, *Trans. Faraday Soc.*, **31**, 1051.

- FINCH, G. I., and WILMAN, H., 1936, *Proc. Roy. Soc. A*, **155**, 345; 1937 a, *Trans. Faraday Soc.*, **33**, 1435; 1937 b, *Ergebn. exakt. Naturw.*, **16**, 353.
KIKUCHI, S., 1928, *Jap. J. Phys.*, **5**, 83.
KIRCHNER, F., 1932, *Ergebn. exakt. Naturw.*, **11**, 64.
SHINOHARA, K., 1932, *Sci. Pap. Inst. Phys. Chem. Res., Tokyo*, **18**, 223, 315, **20**, 39; 1935, *Phys. Rev.*, **47**, 730.
SHINOHARA, K., and MATUKAWA, K., 1933, *Sci. Rep. Inst. Phys. Chem. Res., Tokyo*, **21**, 21.
THIRSK, H. R., and WHITMORE, E. J., 1940, *Trans. Faraday Soc.*, **36**, 565.
WILMAN, H., 1948, *Proc. Phys. Soc.*, **60**, 341.
WYCKOFF, R. W. G., 1931, *The Structure of Crystals* (New York: Chemical Catalog Co.).

Electron Diffraction from Anthracene with Special Reference to the Diffuse Background*

By J. SKREBOWSKI

Physics Department, Imperial College †

MS. received 25 June 1947; read at Science Meeting, 31 October 1947

ABSTRACT. The object of the research was to examine the diffuse pattern obtained when an electron beam is transmitted through a crystal of anthracene. A special camera adapted for low temperature work is described. It has been found that the diffuse pattern has a fine structure which is more apparent in the case of thick crystals. The changes in intensity of the diffuse pattern with temperature have been established experimentally and it is shown that the phenomenon of the diffuse scattering is analogous to that found in x rays.

§ 1. INTRODUCTION

ELECTRON diffraction transmission photographs from single crystals of anthracene show, besides a cross-grating pattern, a superimposed pattern which appears to be that of diffuse bands and areas. These bands and areas are related to the orientation of the anthracene molecule within the crystal lattice and to the separation of carbon atoms in the anthracene molecule; this led Charlesby, Finch and Wilman (1939) to the assumption that the diffuse areas are due to what they call "molecular scattering".

The following is a brief summary of their theory. The molecules are vibrating independently of each other in the crystal lattice and their vibrations have no definite phase relationship with each other. Although built in the crystal lattice they are producing an effect similar to that of a gaseous mass of oriented molecules in the electron beam. In the consideration of the vibrations the intramolecular heat motion is neglected. Molecules are thus vibrating as rigid units about their mean positions in the crystal lattice. As the anthracene unit cell contains two molecules in different orientations the authors proceed to construct qualitatively a molecular scattering pattern such as would be given by two independent molecules in these orientations. This diffuse pattern is to be superimposed on the cross-grating pattern of the crystal. It is claimed that the calculated diffuse pattern agrees well with the observed positions of diffuse areas and bands and gives the correct intensity distribution. Furthermore,

* Part of Thesis submitted to London University for Ph.D. degree examination.

† Now at Imperial Chemical Industries, Billingham.

they have derived an expression for the dependence on temperature of the intensity of the molecular pattern, on the assumption that the thermal displacement of any molecule from its mean position is given by the Maxwell-Boltzmann law.

No confirmation for either the theory or the experiments of Charlesby, Finch and Wilman has been found in recent x-ray work. In this field another temperature-dependent phenomenon was found. It was examined experimentally by Laval, Preston, Lonsdale and Smith, and others, and it received theoretical treatment by Faxen, Waller, Born, Zachariasen *et al.* (see review by Lonsdale 1943). The experimental researches of Lonsdale and Smith (1941) were carried out on organic as well as inorganic crystals. No x-ray investigations have been reported on anthracene, but benzil, sorbic acid, hexamethyl-benzene, urea nitrate and urea oxalate have been studied. A very brief outline of the work of Lonsdale and Smith will be given here.

A simple theory of diffraction of monochromatic x-rays by a crystal lattice predicts that with most crystal orientations (where the Ewald sphere of reflection misses all the reciprocal lattice points) no trace of any pattern should be found on photographs, no matter how long the crystal is exposed to x-rays. Experimentally, however, it was found that, with long exposures or very powerful x-ray sources, some definite pattern of diffuse spots of various shapes, and streaks of different types (sometimes connecting the diffuse spots), could be seen on the photographic plate. The effect was specially clear when highly monochromatized radiation was used. The intensity of the diffuse pattern was greatly diminished when the specimen was cooled with liquid air. The temperature change was reversible.

The diffuse pattern varied considerably from substance to substance and showed very interesting features for some chain and layer structures—for chain structures a characteristic sharp streak and for layer structures a large diffuse spot. The diffuse streaks connecting diffuse spots are found to lie along or nearly along some reciprocal lattice axes. Generally the intensity of the diffuse spot is largest near the position of a Bragg spot, and although the intensity diminishes with crystal rotation from this position, diffuse reflections could sometimes be observed when the crystal was turned as much as 30° . The intensity of a diffuse spot is related to the intensity of the corresponding Bragg reflection; thus no diffuse spots were observed corresponding to reflections forbidden by structure-factor considerations.

The phenomenon of diffuse scattering is accounted for by the Faxen-Waller theory which not only predicts its existence but also gives the formulae for the shape of the diffuse spots and accounts for the dependence of their intensity on temperature. The Faxen-Waller theory assumes that atoms are vibrating thermally in the crystal lattice and that the vibrations are harmonic. As the maximum frequency of atomic vibrations ($\sim 10^{13}$ sec.⁻¹) is much lower than that of x-rays (10^{18} sec.⁻¹) the atoms or molecules in a crystal will appear stationary as far as the passage of x-rays is concerned but slightly displaced from their mean positions. It is possible to consider the atomic displacement as due to superimposed, standing, elastic waves in the crystal which can be of acoustical or optical type. Although the complete theory must take into account both acoustical and optical branches of the vibrational spectrum, nevertheless the acoustical, which corresponds to larger amplitudes, is of primary importance. The theory shows that a single stationary wave of this type would produce two

extra points very near each main reciprocal lattice point. The shorter the wavelength the larger the distance of these extra spots from the main spot. The effect of thermal motion can be represented by a cloud of such extra points, giving extra reflecting power (or diffuse reflection) to every region surrounding a reciprocal lattice point. The shape of this extra reflecting cloud is dependent, among other factors, on the elastic constants (Lonsdale 1943). So far, the Faxen-Waller theory has not been worked out in detail for organic crystals but it explains qualitatively the peculiar diffuse reflections from chain-like molecules and layer structures and also the diffuse streaks extending between two spots.

The results predicted by the molecular scattering theory of Charlesby, Finch and Wilman, and those predicted by Faxen and Waller are essentially different. While, according to the molecular scattering theory, it should only be possible to observe relatively large diffuse areas or bands without any associated fine structure, the Faxen-Waller theory predicts diffuse spots and streaks. Later, Charlesby and Wilman (1942), on re-examining their plates, found evidence of diffuse streaks as well as diffuse areas on the electron diffraction pattern of anthracene. There is also an attempt by Charlesby (1942) to reconcile the molecular scattering theory with that of Faxen-Waller, on theoretical grounds.

The aim of the present research is to study the diffuse areas observed by Charlesby, Finch and Wilman, and to see whether these can be interpreted in terms of the Faxen-Waller theory which has been used successfully in the x-ray work.

§ 2. APPARATUS

A camera of the type designed by Thomson and Fraser (1930) was modified so as to accommodate four photographic plates. This design of camera uses a cold-cathode discharge as electron source and a fine beam is obtained by pin-hole collimating, no magnetic focusing being employed. The accelerating voltage of the electrons was usually in the region of 28 kv. The distance from specimen to photographic plate was 42 cm.

As the experiments were to be carried out at low temperatures, the main cone was redesigned to enable the specimen to be cooled down to the temperature of liquid air (figure 1). A large brass cone B is bored out. Inside it a tube T, about 3 cm. in diameter and 25 cm. in length, made of German silver, is soldered at its upper end to a flexible bellows C, which in turn is soldered to the main cone. The position of the tube can be adjusted by three screws. The tube T at its lower end is closed by a brass disc D. A part of the specimen holder H is tightly screwed into D so as to ensure good thermal contact between these two pieces. The specimen holder, made of copper, consists of two thick plates H and F, both with hemispherical cavities. These are put together with four screws. A thick ring with a rod-like extension A, on either side in the plane of the ring, is enclosed between the plates and can be rotated about its axis. After H, A and F are put together, a short cylinder R can be inserted into the ring from outside and, as the cylinder R fits tightly into the ring, good contact is ensured between them. The specimen-carrying cylinder R is a piece of thick copper tube with a fine wire gauze soldered over the top. On this gauze a transmission specimen can be placed. In the main cone there is a small cone M which can be rotated about its axis. To its lower end a violin string is attached and also fixed to a small wheel W at the end of the axis A. Thus by rotating cone M, A can

also be rotated. With these arrangements it is possible to ensure all the necessary movements, viz. up and down by three screws, rotational movement of the whole cone B, and rotational movement of the specimen-cylinder R. A calibrated copper-constantan thermocouple is soldered to H. All the temperature measurements were made with a potentiometer bridge.

After the camera was evacuated, liquid air was introduced into the tube T and it generally took 5–10 minutes to cool the specimen. The amount of liquid air necessary to keep the specimen at low temperature is small. In the initial experiments at low temperature it was found that the pattern observed on the fluorescent screen became blurred after a few minutes. It was probably due to condensation of organic vapour from the diffusion pump on cold crystal surfaces and it was necessary to introduce on both sides, H and F of the specimen holder, two shields Z in the form of tubes about 2.5 cm. long. After this modification the crystal could be kept for a long time without any appreciable change in the diffraction pattern.

It should be noted, however, that as soon as the apparatus is cooled, a thermal contraction takes place and removes the observed crystal from the beam. Although it should be possible with screws to bring the specimen holder back to its previous position, in practice it was found impossible to find a very small crystal again once it was lost from the beam. A slightly different arrangement was therefore tried. Two pin holes, S_1 and S_2 , were fitted into the shielding tube on the side from which the beam is incident, so that the nearest hole was only about 0.5 cm. from the carrier R. The distance between the holes was about 3 cm. and their diameter was 0.1 mm. With this arrangement it was possible to find the same crystal after lowering its temperature. It also produced a very fine beam and reduced the area of the crystal upon which the beam was falling. The experiments with this device proved particularly successful in elucidating the problem of the fine structure of the diffuse areas.

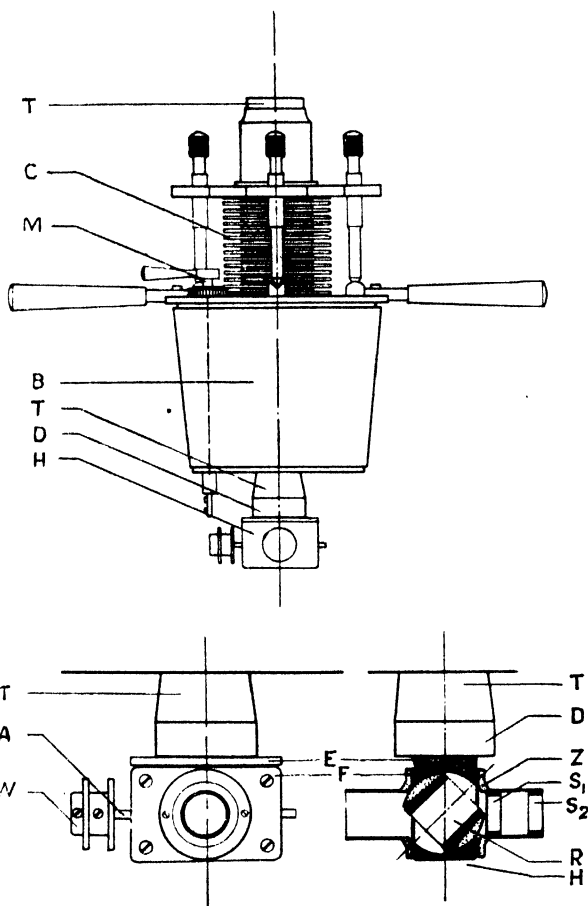


Figure 1. Specimen holder arrangements for orienting specimen and cooling device.

§ 3. CRYSTALLOGRAPHIC DATA. PREPARATION AND MICROSCOPIC EXAMINATION OF ANTHRACENE CRYSTALS

Anthracene ($C_{14}H_{10}$) has a melting point of $217^{\circ}C$. and sublimes slowly *in vacuo* at room temperature. It belongs to the monoclinic system, the elementary cell having the following dimensions:— $a=8.58\text{ \AA}$., $b=6.02\text{ \AA}$., $c=11.18\text{ \AA}$., with $\beta=125^{\circ}$. Anthracene crystals occur in the form of hexagonal flakes, the plane of the flake being (001); they can easily form twins. The unit cell contains two molecules, the atomic coordinates of which were obtained by Robertson (1933). As anthracene crystals are often contaminated with naphthacene, very pure material was used in order to eliminate the possibility of the presence of foreign molecules which might affect the diffuse areas. It was found that this highly purified anthracene formed better crystals than the commercial material.

Thin crystals of anthracene suitable for electron diffraction were prepared by heating a sample of the pure substance in a test tube. A large number of minute crystals showing interference colours was formed from the vapour and as they floated in the air they were easily picked up on a fine gauze. Another way of preparing suitable crystals was to evaporate xylene solution very slowly on a very thin collodion film. Crystals formed by each of these methods were examined and photographed under the microscope. Their appearance differed according to the method of preparation.

While crystals grown from xylene solution formed regular flakes lying sometimes on top of each other, those grown from vapour often formed clusters, twins or intergrowths at different angles. A general view of crystals grown from vapour on to a glass slide as they appear under the microscope is shown in plate 1. During observation under the microscope the crystals disintegrated by developing several holes, although the flake itself was still thick. It seems plausible that the electron beam covers portions of the crystal of very different thicknesses.

§ 4. GENERAL DESCRIPTION OF THE SPOT PATTERN OF ELECTRON DIFFRACTION PHOTOGRAPHS

Most of the electron diffraction photographs obtained by transmission through anthracene crystal flakes showed a fairly extensive pattern due to the *ab* cross-grating inclined to the beam at different angles. Sometimes the pattern was very extensive. This corresponded to a very thin crystal, possibly only a few molecules thick. Other photographs showed only a few intense spots and were correspondingly due to much thicker crystals. The pattern on the photographs was mainly rectangular, or nearly so. Most photographs, whether corresponding to thick or thin crystals showed irregularities in the cross-grating array of spots, frequent doubling or tripling of the spots, convergence of some rows, and generally untidy and irregular grouping of spots. From a large number of photographs taken with crystals grown from vapour, only a few showed a perfect cross-grating pattern such as would be expected from a single crystal; even on these a very noticeable regular doubling of spots was observed. Crystals grown from xylene solution on to collodion did not show these double spots, but unfortunately the rings due to the collodion film tended to obscure considerably the details of the anthracene pattern, especially near the centre.

It is of interest to note that spots forbidden by the structure factor, of the type $(2n+1, 0, m)$, $(0, 2n+1, 0)$, where n and m are integers, appeared on all photographs. They were rather faint and quite sharp.

§ 5. PATTERN INDEXING AND ESTIMATION OF CRYSTAL THICKNESS

The plates were indexed by assuming the values for the cell dimensions given by Robertson (1933). Sections of the reciprocal lattice were drawn to scale and to allow for the crystal being thin, lines were added to every lattice point in a direction perpendicular to the surface of the crystal. This was taken to be an (001) face. The sphere of reflection was then orientated to give the observed pattern.

Crystal thickness can be estimated from the length of the extension of the reciprocal lattice points. If some spots fail to appear on the photographic plate then the reflection sphere misses the corresponding (extended) reciprocal lattice point because its extension is too short. Once the length of the extension is found the crystal thickness is estimated from the formula: $l = 1/t$, where t is the crystal thickness and l is the length of the extension measured from the centre. It may be remarked that the reciprocal lattice points of anthracene are fairly close together. Hence, with crystal thickness below 20 Å., the reciprocal lattice points are so much extended that they form continuous lines. Thus it is not possible to estimate crystal thickness below that value.

§ 6. DESCRIPTION OF THE DIFFUSE PATTERNS

(a) *Experimental results at room temperature*

The character of the diffuse background in the diffraction patterns appeared at first to depend on the crystal thickness and the plates were accordingly classified in this way. The deductions below were made from a large number of plates of which examples, illustrating the salient points, are reproduced.

(i) *Thick crystals* ($t > 200$ Å.). Thick crystal patterns are characterized by relatively few Laue spots, most of which are intense. These patterns show, on well exposed photographs, an array of diffuse spots and streaks, but nothing which can be termed a diffuse area. Plate 2 shows conclusively that the diffuse background has a well defined structure, the details of which can be seen in the enlarged portion (plate 3). These bear a striking likeness to the diffuse background of benzil, in plate 10 *b* of the paper by Lonsdale and Smith (1941). To obtain detail, it is particularly important to use plates which show good contrast (e.g. photomechanical plates show more detail than process plates). It is also necessary to use a fine electron beam to avoid the pattern being a composite one of a number of crystals. In these composite patterns the superposed diffuse background may have the appearance of a diffuse area, as the detailed structure tends to be blurred. That the diffuse pattern is very faint, can be seen by comparing its intensity with that of Laue spots visible in the same photograph (plate 3). The dependence of the pattern on temperature will be discussed below.

Another feature which has been noticed on a large number of plates is that some of the diffuse spots are elongated and not circular. In special cases these are very elongated and are found on both sides of the central beam. Plate 4 shows a typical example of this. These elongated spots were fairly persistent when the orientation of the crystal was changed.

(ii) *Thin crystals* ($t \sim 100$ Å.). Thin crystal patterns are characterized by an extensive array of cross-grating spots. These spots vary a great deal in intensity (roughly in accordance with the structure factor) and forbidden spots are also present. A typical example of the thin crystal pattern is shown in plate 5:

this shows also the type of background called diffuse areas by Charlesby, Finch and Wilman (1939). This is a composite pattern due to a number of crystals in more or less the same orientation. Both the print and the plate show some signs of fine structure, though it is not so evident as for the thick crystals.

In order to eliminate one disturbing factor—the composite nature of the specimen—a special arrangement described in § 2 was introduced which ensured a very fine electron beam. This did not result in perfect single crystal patterns but under the best conditions clearer photographs were made possible. Further, particular care was taken in the time of exposure. Too short exposure gives very little contrast resulting in vague patches of background and a loss of detail. Excessive exposure leads to halation and again loss of detail. The choice of the right exposure time is thus essential.

Plates 6 and 8 are patterns of the same crystal and practically the same orientation taken at two different times (see below) with a fine beam. In the first case, plate 6, the exposure is rather short, and though a fine structure can be seen on the original plate, it is not very evident in the reproduction; in plate 8, on the other hand, the exposure is stronger and the fine structure of the background (for instance the streaking between spots) prominent.

(b) *Experimental results at low temperatures*

The object of the experiments at low temperatures was to find whether the diffuse background disappears or weakens at the temperature of liquid oxygen.

An ideal experiment would be to take a photograph at room temperature and then, with unchanged crystal setting and the same beam intensity, to cool the crystal down to the temperature of liquid oxygen and take another photograph. This was, in fact, done by Lonsdale and Smith with organic crystals using x rays. Unfortunately, in electron diffraction with the present experimental arrangement, it proved to be impossible, as neither the setting of the crystal nor the beam intensity could be strictly controlled. The procedure adopted was to take a photograph at room temperature, cool the specimen-holder to the temperature of liquid oxygen, find the same crystal and take several photographs with different exposures but using the same voltage. It was difficult to find two photographs which were exactly comparable; very slight movement of a specimen could change the spot pattern, though the new pattern could be identified as being due to the same crystal with slightly different orientation or a slightly different thickness.

Plate 6 taken at room temperature shows diffuse patches which, on the original plate, have a fine structure consisting of diffuse streaks joining cross-grating spots. The pattern indicates that the crystal is thin. Plate 7 shows the same crystal cooled to the temperature of liquid oxygen. Although the exposure is stronger than that of plate 6 very little diffuse streaking is visible on the original plate and the strength of the whole background is considerably reduced. The cross-grating spots are also stronger than in plate 6. Plate 8 shows the same crystal warmed to room temperature again; the background is more prominent than in plate 7, and the fine structure more obvious than in plate 6. These three photographs show that the temperature phenomenon is reversible. A number of sets of photographs of the above type were taken, all of which showed the same features.

§ 7. DISCUSSION

The existence of diffuse spots and streaks, observed with anthracene crystals, suggests that the effect is the same as that found by Lonsdale and Smith (1941) in their investigation on the x-ray diffuse scattering from some organic crystals. As mentioned above, the x-ray diffuse pattern of benzil (Lonsdale and Smith 1941, plate 10 *b*) bears a striking resemblance to the pattern shown in plate 3. In other photographs, e.g. plate 4, the characteristic elongated streaks on either side of the central beam, and their observed persistence with tilt of the crystal are similar to the streaks observed in x rays with sorbic acid (Lonsdale and Smith 1941, plates 12 *b* and 21 *a, b, c*).

The temperature effect cannot be demonstrated as clearly as with x rays because of considerable experimental difficulties, but the examination of several sets of photographs, e.g. plates 6, 7, 8, shows conclusively that the diffuse pattern intensity diminishes considerably at the temperature of liquid oxygen. This decrease is especially well marked in the case of diffuse streaks joining the diffuse spots; the streaks disappear almost completely at low temperature. Thus it can be concluded that the effect of diffuse scattering as found in x rays has a close analogy in electron diffraction.

The diffuse scattering in x rays was interpreted successfully in terms of the crystal vibrations and it is reasonable to conclude that a similar phenomenon in electron diffraction is due to the same cause. It is, therefore, unnecessary to assume that the diffuse spots and streaks are due to the molecules vibrating independently, which has been postulated by Charlesby, Finch and Wilman (1939); indeed, the fine structure observed, plates 2, 3, 4 cannot be reconciled with their theory.

Emphasis has been placed by Charlesby, Finch and Wilman (1939) on the fact that the diffuse patches observed on their patterns from anthracene crystals are due to independent scattering from anthracene molecules, and that, therefore, the molecular structure of anthracene can be deduced from the relative positions of the diffuse areas on the patterns. While it is possible that some molecules at the surfaces and edges of the solid may scatter independently, the molecules forming the lattice do not.

Examination of the patterns obtained in the present work showed that in general, the diffuse background was strongest in those positions where Laue spot intensities were strongest; this is to be expected on the Faxen-Waller theory. It can be shown on this theory (Born 1942-43) that for any crystal the intensity of the diffuse scattering is directly proportional to the square of the lattice structure factor, in the region near the Laue spot. One would therefore expect that an examination of the structure factor, F , would show where the diffuse background would be most intense. Contour lines of F^2 , for anthracene, corresponding to the orientation in plate 5, have been drawn, and are shown in figure 2. The positions of the strong regions coincide reasonably well.

On the molecular scattering theory, the intense regions of the background should be given by each molecule scattering independently. This results in a slightly different set of contour lines from the previous set (see figure 3) but they are not significantly better as a rough representation of the intensity distribution found on the plate.

The orientation chosen in plate 5 is a fair one from the point of view of the molecular scattering theory, since the cross-grating spots are grouped, as regards

intensity, into only three regions. Neither set of contour lines (figures 2 and 3) shows any detailed structure over a region comparable with the cross-grating unit. A careful examination of plate 5, however, shows that there is a streaking between spots and a fine structure. The molecular scattering theory fails, therefore, to provide a basis for a full description of the experimental results.

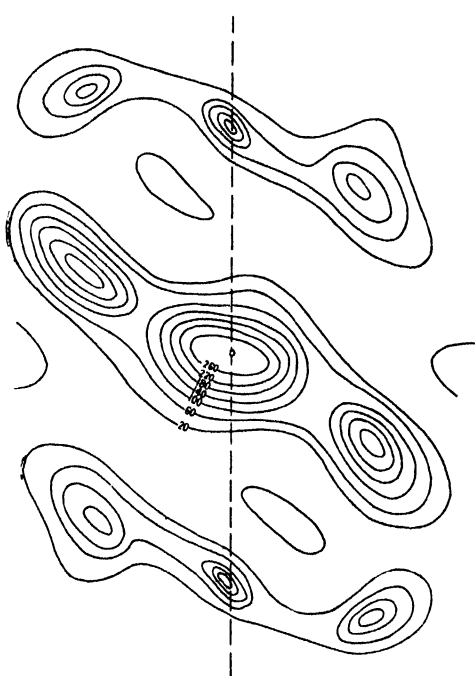


Figure 2. Contour lines of intensity scattered from a unit cell of anthracene. Beam direction as in plate 5.

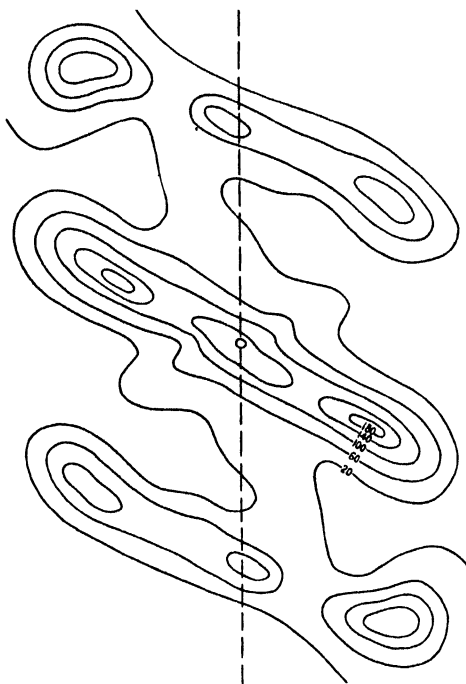


Figure 3. Contour lines of intensity scattered from two independent molecules oriented as in the anthracene cell. Beam direction as in plate 5.

ACKNOWLEDGMENTS

The author wishes to express gratitude to Professor Sir George Thomson, F.R.S., for providing facilities for the research, and for his great help and interest in these investigations, to Dr. M. Blackman and Dr. W. Cochrane for their valuable help and interest throughout the whole period of the research. He also wishes to thank Mr. B. K. Johnson of the Technical Optics Department of the Imperial College of Science and Technology for the microscopic photographs, Mr. L. J. Bowen, F.R.S., of the Physical Chemistry Laboratory, Oxford, and Dr. Schnurmann and Dr. Boyd-Barrett of the Manchester Oil Refinery Ltd. for examples of pure anthracene.

REFERENCES

- BORN, M., 1942-43, *Rep. Prog. Phys.*, **9**, 294.
- CHARLESBY, A., 1942, *Proc. Phys. Soc.*, **54**, 379.
- CHARLESBY, A., FINCH, G. I., and WILMAN, H., 1939, *Proc. Phys. Soc.*, **51**, 479.
- CHARLESBY, A., and WILMAN, H., 1942, *Nature, Lond.*, **149**, 411.
- LONSDALE, K., 1942-43, *Rep. Prog. Phys.*, **9**, 256.
- LONSDALE, K., and SMITH, H., 1941, *Proc. Roy. Soc. A*, **179**, 8.
- ROBERTSON, J. M., 1933, *Proc. Roy. Soc. A*, **140**, 79.
- THOMSON, G. P., and FRASER, C. G., 1930., *Proc. Roy. Soc. A*, **128**, 641.

DISCUSSION

Dr. K. LONSDALE. Mr. Skrebowski's paper is of great interest, in that it suggests that experimental data on diffuse scattering, which appeared to have been satisfactorily accounted for by the assumption of independently vibrating molecules, can be explained equally well by assuming that the diffuse intensity is dependent upon the ordinary structure factor, being greatest in the neighbourhood of reciprocal lattice points with the largest reflecting power. That this is true in general is quite clear to anyone who has worked in this field; it may be seen illustrated in plates 5 and 6 of Lonsdale and Smith, *Proc. Roy. Soc. A*, 1941, 179, 8, for x ray diffuse scattering; although of course the rule cannot be applied in detail for all structures.

It is interesting also that the photograph (author's plate 7) taken at liquid air temperatures shows very clearly not only the decrease of intensity of the diffuse pattern but also the increase of intensity of the Laue spots (especially of those at large diffraction angles), which is a necessary corollary. This fact has misled Dr. Wilman into thinking that plate 7 must have been taken using a thinner part of the crystal than that which gave plates 6 and 8.

Another point which comes out more clearly perhaps from the electron diffraction photographs than from x rays (although quantitative ionization spectrometer measurements given by Laval for KCl, and Geiger counter measurements on urea and on oxalic acid dihydrate shortly to be published in the *Acta Crystallographica*, do prove this fact without any doubt) is that the falling off of diffuse intensity with increasing angle of diffraction is markedly less than that of the Laue intensities. This also follows from the Faxen-Waller theory, as extended by Born and others.

I see no particular significance, however, so far as this controversy on interpretation is concerned, in the presence or absence of diffuse "areas", meaning by these the overall blackening of considerable regions of the photographic plate. The practical applications of the Faxen-Waller theory so far made have only considered the parts of reciprocal space in the neighbourhood of reciprocal lattice points, but in all but the simplest cases (and anthracene is not a simple case) it is quite certain that optical waves will have to be taken into account, and that the effect of these will be greatest *between* the reciprocal lattice points. In one sense the effect of these optical vibrations is likely to be much more like that of independent molecular vibrations than the acoustical vibration effects would be, but I am not clear that the present method really provides a crucial experiment by means of which a distinction can be made. The presence or absence of diffuse scattering in the neighbourhood of reciprocal lattice points of zero Laue intensity is not such a criterion, for any vibration or distortion which disturbs the crystal lattice will necessarily also disturb the conditions which cause the Laue reflection to disappear. The surprising thing, perhaps, is that such "forbidden" reflections do not appear more often in diffuse form. Some unpublished work by Mr. P. G. Owston on the diffuse pattern of ice shows that such forbidden reflections do appear in that case, where the intermolecular forces are much stronger than they are in anthracene, and independent molecular vibrations are very unlikely.

My opinion is that the electron diffraction photographs of thick anthracene crystals are the most useful for comparison with x-ray diffraction effects, because the use of very thin crystals will necessarily introduce a possible confusion between surface and internal structure, and because with very thin crystals some of the long wavelength standing waves will no longer exist.

Professor Sir GEORGE THOMSON. This is a paper which raises an interesting theoretical issue. It is one of those cases in which two widely different theories lead to nearly the same answer. Born's theory has been well confirmed for x rays by the work of Dr. Lonsdale and others, and there can be no doubt of its essential truth. Unfortunately it is very complicated, and involves a number of elastic constants. For substances as complicated and crystal forms as unsymmetrical as anthracene, it is not possible to work it out in detail. One is therefore driven to qualitative predictions, as, for example, that the diffuse effects are most marked near points of the reciprocal lattice with large structure factor. It is inherently extremely probable that the Born effects will also occur with electron diffraction, and I am convinced that the effects first observed by Finch and Wilman and studied in this paper are the electron analogue of those studied for x rays by Dr. Lonsdale. The general agreement in appearance, the close correlation with structure factor, the character of the fine

structure observed by Dr. Skrebowski, all confirm this. But that is not to say that the theory put forward by Charlesby, Finch and Wilman, which ascribes the effect to diffraction from molecules vibrating independently under thermal agitation, is necessarily wrong. The Born theory is a theory, and I believe a complete one, of the effect of thermal agitation on the scattering of x rays. It is expressed in terms of certain elastic constants. These elastic constants are determined by the structure of the crystal and the forces between the atoms in it. If certain groups of atoms are held together very strongly so as to form molecules, and if the forces between molecules are relatively weak, these facts will be reflected in the values of the elastic constants, and by the time you have gone through all the mathematics it is probable you will get roughly the same answer as if you had started by assuming the molecules to be independently moving rigid solids as Charlesby, Finch and Wilman have done. Only you will not get the fine structure which only the more refined theory is capable of predicting. I regard the independent rigid motion of the molecules as inherently probable though not certain; a flat molecule like anthracene might easily bend, but I do not think that experiments of this kind are capable of proving it, since there is so little difference, at least in the cases tried so far, between the distribution of intensity deduced from the structure factor and that deduced from the hypothesis of independent molecular motion.

Dr. A. CHARLESBY. The problem with which Mr. Skrebowski deals, i.e. the presence of diffuse bands in electron diffraction, is of considerable value in the determination of molecular structure. Under certain conditions it enables the arrangement of atoms in the molecule to be determined directly, without going through an elaborate Fourier analysis of Laue spot intensities. Several papers have shown their occurrence and their correlation with molecular structure.

Mr. Skrebowski attempts to explain away the existence of these diffuse bands by ascribing them to the blurring or smearing out of relatively sharp streaks. These sharp streaks are, however, quite distinct and have indeed been previously observed in electron diffraction and x-ray work. I fail to see how photographic plates can smear out a streak perhaps 1 mm. wide into a band 5 mm. in width, while yet reproducing sharp spots less than 0.1 mm. across.

Using Process plates, the author was unable to obtain diffuse bands, but observed sharp streaks. This does not disprove the existence of these diffuse bands. His observation arises simply from the sharp contrast and limited intensity range of these plates, which suppress low intensities. It is for this reason that after experimenting, we preferred Special Rapid to Process plates. Obviously it is not a sufficient explanation to show that under unsuitable conditions of exposure a phenomenon does not appear. It is far more important to explain why it does appear under other conditions.

I would like to refer to another question raised in the paper, namely the comparison of diffuse streaks and diffuse bands. On the basis of rather superficial similarities, mainly that both phenomena are more diffuse than the spot pattern, Mr. Skrebowski draws the conclusion that bands and streaks are identical, whereas a closer examination reveals wide differences between them.

My conclusion is that we are dealing with three distinct sets of phenomena: (a) sharp spots corresponding to the crystal structure; (b) diffuse streaks and spots, observed both in x-ray and electron diffraction work; these streaks arise from thermal motion and correspond to elastic waves throughout the crystal; (c) diffuse bands, so far only observed in electron diffraction, whose non-appearance in present-day x-ray work is readily explained (Charlesby, *Proc. Phys. Soc.*, 1942, **54**, 379).

In the case of a few simple structures, the effect on the diffraction pattern of elastic vibrations throughout the crystal may be evaluated. These vibrations give rise to streaks and spots, the presence of which is confirmed by x rays. In the case of more complex crystals the mathematical difficulties appear insuperable, and one is obliged to proceed by qualitative comparisons. Similar streaks and diffuse spots have been observed in electron diffraction patterns and reported in *Nature*. It may be presumed that they arise from the same continuous type of thermal vibration.

Instead of considering these continuous thermal vibrations throughout the crystal, one can go to the opposite extreme, and consider each molecule as vibrating independently of its neighbours. On this alternative assumption, the mathematical calculation can be readily completed even for a complex crystal such as anthracene. There is very good

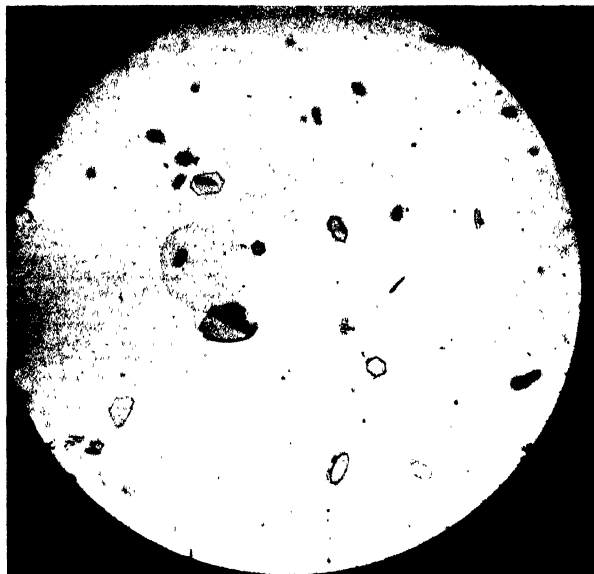


Plate 1. Anthracene crystals grown from vapour. Magnified $\times 50$.



Plate 4. Thick crystal pattern of anthracene showing fine structure and two elongated diffuse streaks.

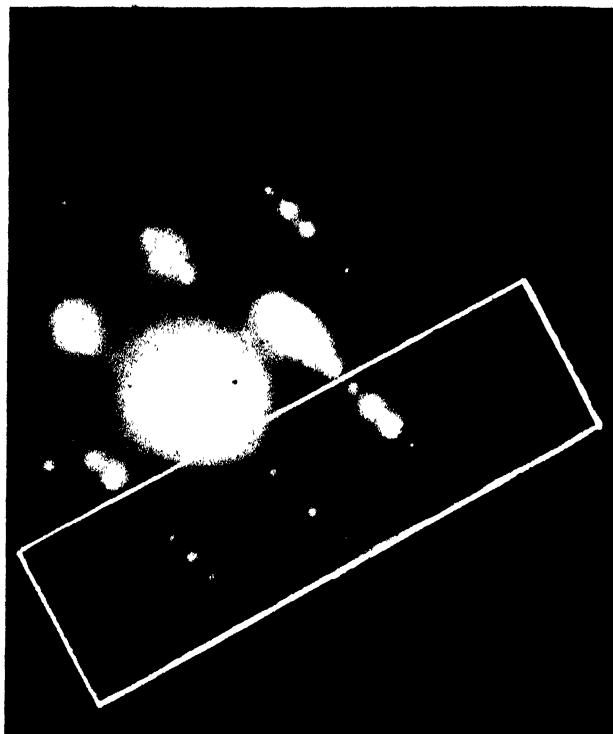


Plate 2. Thick crystal pattern of anthracene showing fine structure



Plate 3. Enlargement of a portion marked on photograph 2, showing details of the fine structure.

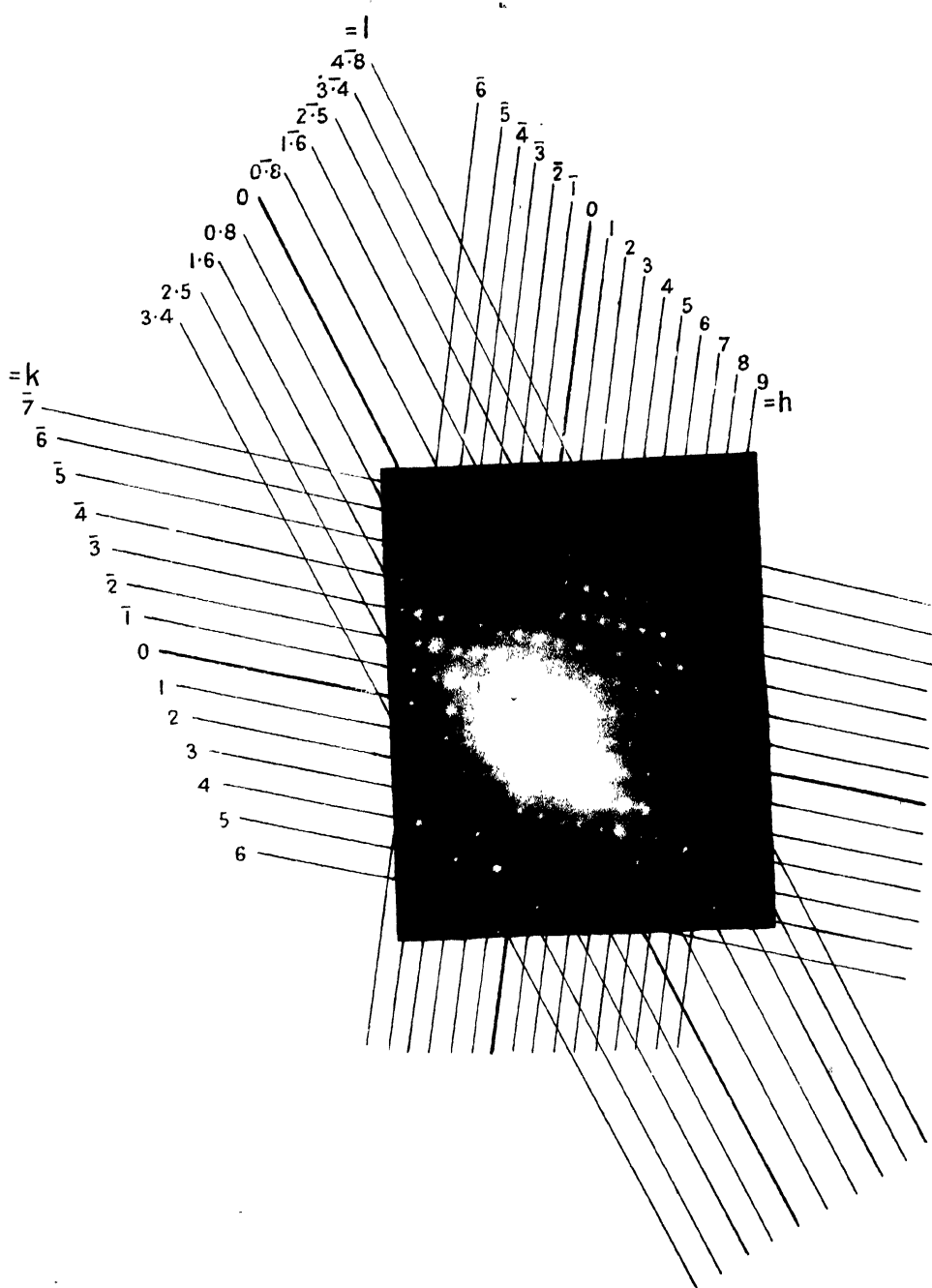


Plate 5. Indexed pattern of a thin crystal of anthracene showing diffuse background. Incident beam making angles with axes a , b , c , $\theta_a=107^\circ$, $\theta_b=105^\circ$, $\theta_c=18^\circ$.

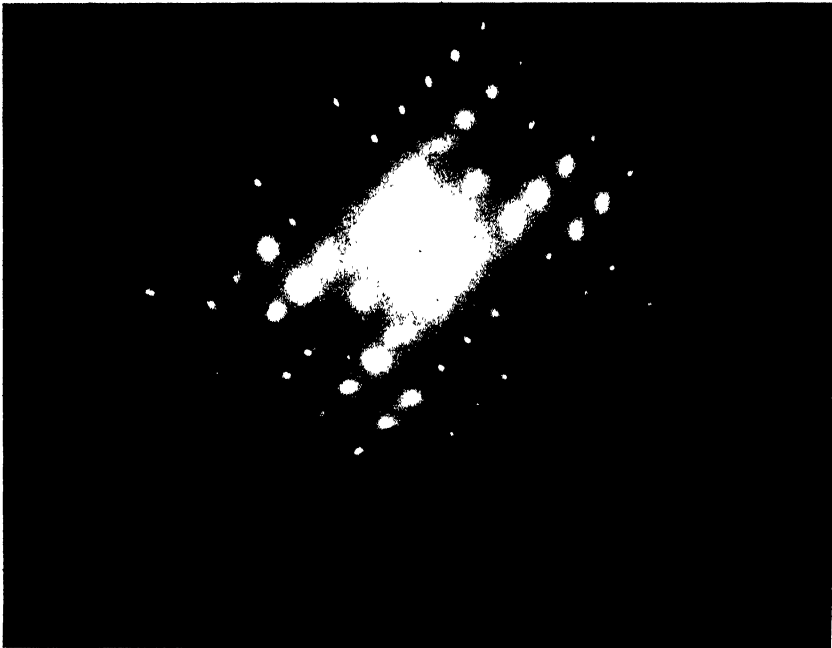


Plate B.

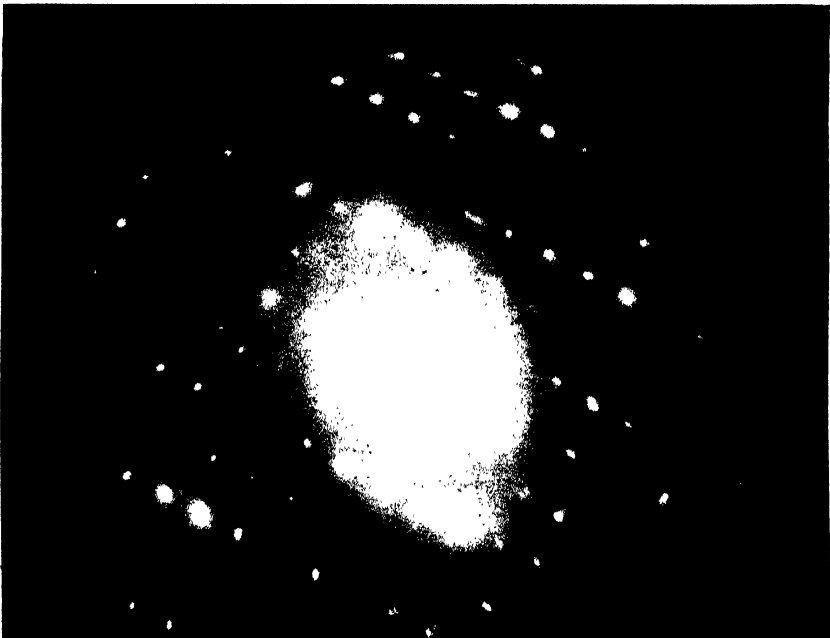


Plate C.

agreement between the calculated pattern and the diffuse bands observed. This theory was put forward in 1939 by Charlesby, Finch and Wilman. It was realized that this explanation was only a first approximation, since it did not account for the presence of diffuse streaks and spots, which could often be observed superimposed on the diffuse bands.

It appears to me that both theories are necessary to account for the full pattern. Of the total thermal motion of a molecule one component can be considered to form the elastic waves continuous throughout at least fairly large areas of the crystal. This component leads to streaks and diffuse spots. The remaining component is independent of that of other molecules and gives rise to the diffuse bands.¹ But I must emphasize that an independent component must be accepted to account for the diffuse bands.

Mr. Skrebowski assumes that for a complex crystal such as anthracene the intensity of diffuse streaks follows the structure factor of the unit cell. No such calculation has been made, nor does the diffuse pattern observed in simpler crystals such as diamond follow this intensity distribution.

In his remarks the author points out that the structure factor for the unit cell is not very different from that calculated from each molecule separately, and attempts to substantiate this statement with a direct comparison between the two. Unfortunately the single orientation he has shown is very badly chosen to make this comparison. In other orientations the difference is much greater, and in such orientations the observed diffuse pattern does not agree with that calculated from the structure factor, but rather with that calculated for each molecule separately.

Anthracene provides a good example of this distinction. In certain directions the amplitudes scattered by the two molecules are equal but opposite in phase so that the resultant intensity is zero. Spots should therefore vanish (forbidden reflections), as should any diffuse streaks in the immediate neighbourhood. Observations of the area where diffuse patterns from different molecules overlap show no such phenomenon of interference. Strong diffuse bands occur in areas where the interference between the two molecules in the unit cell should eliminate them, and does in fact eliminate the Laue spots. On the other hand the theory of independent molecular scattering accounts both for the appearance and extent of diffuse bands in these areas (plate A).

In view of the importance of this argument, it is worth considering it in further detail. On page 522 of the original paper by Charlesby, Finch and Wilman (*Proc. Phys. Soc.*, 1939, **51**, 479) the difference was pointed out between the two factors in the expressions for the spot and diffuse band intensities.

The expressions for the intensity distribution of the spot and diffuse band pattern respectively in a direction h, k, l were given as

$$I_1 = (A^2 + B^2)\pi(\sin^2 M/l\pi/\sin^2 h\pi) \exp(-16\pi^2 \sin^2 \theta kT/\lambda^2 f),$$

$$I_2 = N \sum_p (A_p^2 + B_p^2) \{1 - \exp(-16\pi^2 \sin^2 \theta kT/\lambda^2 f)\}.$$

A_p, B_p is the amplitude of the wave scattered by molecule p in the unit cell. In the expression for the spot pattern $(A^2 + B^2)$ is equal to $(\sum A_p)^2 + (\sum B_p)^2$ summed over all molecules in the unit cell, the phase difference between different molecules being taken into account. In the corresponding expression for the diffuse bands $\sum (A_p^2 + B_p^2)$ no phase difference between molecules is involved.

In the particular case of anthracene there are two molecules per unit cell p, q . As the molecules are symmetrical $B_p = B_q = 0$. The structure factor for the spot pattern—and according to Mr. Skrebowski for the diffuse streaks—is

$$(A_p + A_q)^2 \text{ or } A_p^2 + A_q^2 + 2A_p A_q \cos \pi(h - k).$$

According to the theory of independent molecular vibration the structure factor for the diffuse bands is $A_p^2 + A_q^2$.

The distinction between these two expressions is best seen in diffractions for which $A_p = A_q$ (eg. $h0l$ and $0k0$). The intensities are then proportional to $4A_p^2 \cos^2 \pi h/2$ or $4A_q^2 \cos^2 \pi k/2$ (spots and diffuse streaks) and $2A_p^2$ (diffuse bands). In the theory of independent molecular scattering the intensity of the diffuse bands is modified by the molecular structure factor and by a thermal factor while that of Laue spots and diffuse streaks in addition depends on the value of the expression $\cos^2 \pi h/2$ or $\cos^2 \pi k/2$.

If the author's thesis is correct, the intensity distribution within a diffuse band which extends over an area corresponding to several Laue diffraction spots should show characteristic fluctuations from $4A_p^2$ to 0. Although certain diffuse bands do show some variations

in intensity, there is no doubt that many bands occur with high intensities at points where the factor $\cos^2 \pi h/2$ or $\cos^2 \pi k/2$ would cause them to disappear. Diffuse bands are often observed at points where the spot pattern is eliminated by the structure factor, e.g. ($h0l$) diffraction with h odd, or ($0k0$) diffraction with k odd. In fact the intensity distribution in the diffuse bands is proportional to the structure factor of each molecule, and not to the structure factor of the unit cell.

One is therefore driven to the conclusion that the diffuse bands although dependent on the atomic arrangement in the molecule are not dependent on the arrangement of the molecules in the unit cell. As has long been suggested—and verified by the author—the appearance of diffuse bands is a temperature-dependent phenomenon. It seems difficult to escape the conclusion that these bands derive from that component of the thermal motion of each molecule which is independent of that of its neighbours.

I have mentioned the existence of a number of characteristic differences between diffuse bands and diffuse streaks. The following observations summarize these differences:

(i) In the patterns shown to us the streaks run from one sharp spot parallel to one of the spot arrays. The bands on the other hand bear no obvious relationship to the direction of the spot pattern. (ii) The intensity of the streaks is modified in the proximity of a diffraction spot. The bands appear to be quite independent of spot intensity except in so far as they both vanish in directions in which the structure amplitude from each molecule is low. (iii) Each streak is relatively narrow. Bands have a width often equal to several times the distance between Laue spots. (iv) The intensity of the streaks, in the photographs shown, is less than that of spots in the neighbourhood. Diffuse bands at some distance from the central spot are far more intense than Laue spots. (v) Although diffuse streaks may be observed superimposed on diffuse bands near the central part of the pattern where they can be readily distinguished, only diffuse bands are observed further away from the centre. (vi) The diffuse streaks are related to the direction of thermal vibrations in the crystal, and hence to the crystalline axes. Their intensity is mainly dependent on the direction of these thermal vibrations. The diffuse bands are independent of crystal orientation except in so far as this determines the molecular orientation.

In conclusion, I would like to state that the assumption of independent vibration leads to a number of deductions which may be readily verified in the diffraction patterns. These include (i) the distribution and dimension of diffuse bands in the diffraction pattern; (ii) the relative intensity distribution within the bands; (iii) the increased intensity of the bands relative to the spot pattern as the angle of diffraction is increased; (iv) the non-occurrence of interference between bands from different molecules.

A simple extension of this idea provides an explanation of the diffuse spots seen by Preston and Bragg in inorganic crystal diffraction patterns.

The theory in the simple form in which it was first presented does not explain the occurrence of diffuse streaks or spots.

The alternative theory of continuous thermal vibrations as propounded by Mr. Skrebowski explains the appearance of diffuse streaks and spots, though only qualitatively. It does not explain the following observations:

- (i) the non-intensification of bands near Laue spots;
- (ii) the non-interference of diffuse bands;
- (iii) the occurrence of strong diffuse bands in areas where the structure factor eliminated the spot pattern;
- (iv) the appearance of diffuse bands at large diffraction angles, when the spot pattern has vanished.

The most likely assumption appears to be that the thermal vibration of each molecule can be split into two components:

- a component dependent on elastic vibrations throughout the crystal and giving rise to diffuse streaks and spots;
- a component which is independent of that of its neighbours, leading to diffuse bands.

From the practical point of view, the latter component is of greater importance in so far as it provides a simple and effective method of molecular analysis.

In view of the increasing interest shown in these extra diffraction phenomena, might I suggest that some less confusing terms be applied to diffuse streaks, diffuse spots and diffuse bands, now that their causes are being separated. Much confusion in this paper arises from the use of a single adjective to describe such diverse phenomena.

Dr. H. WILMAN. Although it seems reasonable, as we have said before, to expect the diffuse maxima to become weaker relative to the Laue spot intensity at low temperature, comparison of plate 7 with 6 and 8 is largely vitiated because plate 7 appears to correspond to a thinner part of the crystal, and the print of plate 7 seems to have had a longer exposure than those of plates 6 and 8. The plate itself still contains considerable intensity of diffuse maxima. This may be due to considerable heating of the crystal by the rather low energy (28 kv.) electrons used. Mr. Skrebowski has thus not conclusively shown the dependence on temperature. I have no doubt, however, but that such a temperature effect must exist.

The apparatus is not new in type. No references to previous work with liquid-air cooled specimens are given, though essentially identical apparatus, avoiding impurities condensing on the specimen, was described by Gen, Zelmanoff and Schalnikoff in 1933, Hass in 1937, 1938, Was in 1939, and Charlesby in 1945 (*Proc. Phys. Soc.*, **57**, 510).

The author suggests that he has discovered pattern details which had, in fact, already been described by Charlesby and Wilman (1942). In plate 5 Mr. Skrebowski has drawn three sets of parallel equidistant straight lines through the spots as the Laue zones—this would mean that the three crystal axes were all co-planar and normal to the beam, which is impossible and inconsistent with the spot intensities even as an approximation. Further, the crystal appears imperfect and the electron beam too broad.

In the introduction and discussion, Mr. Skrebowski expresses views which contradict facts published by Charlesby, Finch and Wilman (1939), and Wilman in 1942 (*Proc. Phys. Soc.*, **51**, 350), Charlesby (1942) and Charlesby and Wilman (1942), which we now illustrate again with slides. The x-ray patterns cannot confirm our electron diffraction results or otherwise, owing to the difference in wavelength, energy, crystal thickness etc. K. Lonsdale (*Proc. Phys. Soc.*, 1942, **54**, 352) has pointed out that the different crystal thickness in the two cases will probably lead to a different distribution of the thermal waves. The theory of thermal-elastic waves in the crystal can account for the diffuse spots in the x-ray patterns and also for the similar spots in the electron diffraction patterns, but has not been able to account for the more diffuse bands and areas of intensity hitherto only observed in the latter. Both the thermal-elastic wave theory and our molecular scattering theory fail to explain the whole of the electron diffraction phenomena, though a mixture of the two does appear to account for the main characteristics (plate A).

Mr. Skrebowski confuses groups of the thermal-elastic spots with the broad diffuse maxima on which they often, but by no means always, lie and which he has not effectively recorded owing to under-exposure and use of plates with unsuitable characteristics. Similarly, the x-ray patterns do not show even the thermal-elastic spots if under-exposed. Our experimental proof of the correspondence of the diffuse maxima with the patterns which would arise from the separate molecules orientated as in the crystal is a fact which he seeks to interpret as due to the thermal-elastic spots being strongest (and the weaker outer parts tending to coalesce) where the normal structure factor is high. He admits his F^2 contour diagram, figure 2, is not significantly different from that for F_1 and F_2 , figure 3, and that neither show subsidiary maxima within the broad diffuse maxima. In most orientations the F^2 distribution *should*, however, in certain regions show maxima and minima separated by distances of the order of that between the Laue spots. This has not so far been reliably observed—we see no appreciable interference of this sort where the maxima of the two differently orientated molecules overlap. Moreover, the diffuse bands and areas are still relatively strong, as expected from the independent-molecule theory, in parts of the pattern distant from the undeflected spot, where the Laue and Bragg spots are no longer strong enough to be visible. Hence, it seems to be the $F_1 + F_2$ diagrams, as we used them, which apply.

Professor G. I. FINCH. In discussing this paper I should like in the first place to outline the history of the discovery of these thermal effects in electron diffraction patterns. The diffuse areas of intensity were first recorded by Finch, Quarrell and Wilman (*Trans. Faraday Soc.*, 1935, **31**, 1051). In 1937, Finch and Wilman (*Ergebn. exakt. Naturwiss.*, 1937, **16**, 353) were able to say that these areas were associated with the molecular structure rather than with the crystal lattice; and in 1938, Finch and Tyson (15th *Ann. Rep. British Emp. Cancer Campaign*, 1938) accounted in a similar manner for the diffuse area patterns they had obtained from a variety of carcinogenic compounds. In 1939, Charlesby, Finch and Wilman (*Proc. Phys. Soc.*, 1939, **51**, 479), made a comprehensive study of the anthracene

pattern and gave detailed proof that the diffuse areas originate from the molecules. In 1942, Wilman (*Proc. Phys. Soc.*, 1942, **54**, 350) published a pattern from an inorganic substance, cadmium iodide, showing diffuse lines along the main spot rows similar to those found with phthalimide (Finch and Wilman 1937) and later, in 1942, Charlesby and Wilman (*Nature, Lond.*, 1942, **149**, 411) pointed out that these diffuse lines correspond to the diffuse spots or lines which had then recently been found in x-ray patterns, and that they occur together with these still more diffuse areas of intensity in electron diffraction patterns obtained from a wide variety of molecular aliphatic and aromatic compounds, including anthracene.

The first clear recording of the coherent thermal scattering effects with x rays was by Preston in July, 1939, some four or five months after the first full publication of our results and four years after we first announced their discovery. It is remarkable that since then the rôle played by electron diffraction in recording thermal scattering effects seems to have been largely overlooked by x-ray workers—I have seen only two references to it, by Mrs. Lonsdale and the late Sir William Bragg.

Thus in 1942, it was clearly shown by us that electron diffraction patterns may exhibit three main types of features, (i) normal Laue spots together with, in thicker crystals, more or less well-developed Kikuchi lines of normal type, (ii) more diffuse spots due to coherent thermal vibrations, like those now so much studied in x rays and, finally, (iii) diffuse areas of intensity due to molecules scattering as independent units as a result of incoherent thermal motion. Mr. Skrebowski, as far as I can see, has added little or nothing to this picture and indeed seems to me to have closed his eyes to the facts which show the occurrence of independent molecular scattering.

I need only draw your attention to the much larger thickness of the crystals used in x rays to show that the absence of these incoherent thermal-motion effects in the x-ray pattern is not surprising. The molecules at or close to the surface of a crystal are obviously subject to unsymmetrical forces resulting from atoms (or molecules) being on each side of them and below, but not above, them in the crystal, so that their motion must be less regular than that of the molecules deeper inside a large crystal; and, moreover, there must be an absence of the longer standing waves in directions steeply inclined to the main surface of a thin crystal sheet. In large crystals there is a small ratio of surface to mass, and the relative importance of the diffraction from the surface layers is small, but with the very thin crystals necessarily used in electron diffraction it will be much larger, and may even predominate in crystals which are only a few molecules thick. This is indeed precisely what is found experimentally, when an adequate technique is employed, not only in production of the pattern, but also in its scrutiny.

I have hinted elsewhere at the potential value of this independent molecular pattern to molecular and crystal-structure analysis. The coherent thermal-motion pattern has been in this respect, as far as I am aware, of only limited value in x-ray analysis. It will therefore, be of some interest if I tell the meeting that Miss Fisher, working in my laboratory, has in proof with the Society*, a paper which describes how an appreciation of the independent molecular scattering pattern has enabled her to make a valuable advance in our knowledge of the molecular and crystal structure of rubber, a field in which hitherto we have had to rely only on x rays as a tool.

AUTHOR'S reply. Some misunderstanding has arisen about part of my paper which I should like to correct. Contour lines of the intensity scattered from a unit cell (figure 2) are drawn for the purpose of indicating where the cross-grating spots (if they appear) are strongest. Thus a cross-grating spot lying between, say, lines 140 and 180 will have an intensity between these values. These contour lines tell nothing about the diffuse background between cross-grating spots. Their sole purpose is to show where strong cross-grating spots will tend to group, as well as the shape of these groupings. Thus Dr. Charlesby's discussion on this part of my paper is inapplicable.

Owing to the length of the discussion I should like to recapitulate briefly the issues involved. Charlesby, Finch and Wilman claim the following:

1. The diffuse areas are primary phenomenon.
2. They are without any detailed structure.
3. They can be explained by the molecular scattering theory which predicts their positions and shapes.

* Now published (*Phys. Soc. Proc.*, 1948, **60**, 99).

4. Sometimes diffuse streaks and spots appear, but these have nothing to do with the diffuse areas. These streaks and spots can be explained in terms of the Faxen-Waller theory and are similar to those found in x rays recently.

In my paper I attempted to show :

1. The diffuse areas (in the Charlesby, Finch and Wilman sense) are secondary phenomena.
2. They always have a definite fine structure which is more prominent in thick crystals.
3. With thin crystals this fine structure is very often smeared out and the appearance of diffuse bands or areas without any fine structure prevails.
4. The diffuse streaks and spots are grouped into bands or areas and the position and shape of these groups is equally well (if not better) predicted by structure factor considerations than by the molecular scattering theory.

The Charlesby, Finch and Wilman view has many conflicting elements. If the diffuse streaks and spots exist independently, then why do their contour lines drawn for the diffuse areas include these streaks (figures 10 *a*, 14 *a*) ?

In the discussion there was no mention of the different appearance in the pattern for thick and thin crystals. This is, however, an important point in the present controversy. The pattern in plate 2 might be taken for the diffuse areas (in the Charlesby, Finch and Wilman sense) but for plate 3 which shows that these areas have a fine structure. If, however, the crystal shown on plate 2 were thinned a large number of cross-grating spots would appear. This would tend to obliterate the fine structure and produce something like that observed by Charlesby, Finch and Wilman (1939), i.e. diffuse areas. In the discussion none of these authors attempted to explain the absence of diffuse areas on plates 2 and 4.

One of the main claims of the molecular scattering theory is that it can predict the position of the diffuse areas satisfactorily. In my paper I have shown that these can be found equally well from the structure factor consideration. They coincide with the regions where the structure factor values are high. This is not a coincidence as Dr. Charlesby suggests, for I have calculated this with widely different crystal orientations. Unless he produces quantitative contour lines for a particular photographic plate his argument will have little value. In fact, the qualitative treatment adopted by Charlesby, Finch and Wilman (1939) is open to criticism. Often their diffuse areas are very small (figures 10 *a*, 14 *a*) and cover 1–2 reciprocal lattice cells. A quantitative treatment would give a different result. The diffuse areas in the latter case are much broader (figure 3) and are without sharply defined limits, in contrast to the Charlesby, Finch and Wilman estimations.

Drs. Charlesby and Wilman in interpreting the diffuse pattern apply either the Faxen-Waller theory or the molecular scattering theory or a mixture of both without proving first that these can be applied simultaneously and are not mutually exclusive.

Experiments have been performed on other molecules in order to test the molecular scattering theory further, for this theory requires large and highly symmetrical molecules to produce the effect (Charlesby 1942). Now phenanthrene is not a symmetrical molecule and its unit cell contains four molecules. If, as the molecular scattering theory requires, each of these molecules were scattering independently, one would expect the resulting pattern to be a maze of diffuse bands and areas. Experimentally, however, the pattern is very simple and highly symmetrical (plate B). It appears to follow the structure factor and is composed of diffuse streaks (visible on the original plate). Naphthalene was the other compound examined. Its crystalline structure is almost a replica of anthracene somewhat reduced. According to the molecular scattering theory, the naphthalene molecule, being much shorter than anthracene, should give very broad diffuse areas. The pattern observed (plate C) has diffuse areas (with fine structure) of the same angular spread as anthracene.

I found Dr. Charlesby's argument about the use of photographic plates in electron diffraction unconvincing.

In replying to Dr. Wilman, I would point out that he has misunderstood the indexing shown on my plate 5. The lines shown are not Laue zones. His comment regarding my apparatus is ill-founded in view of the statement made by his collaborator, Dr. Charlesby (*Proc. Phys. Soc.*, 1942, **54**, 383, § 2), "Experimental difficulties have so far prevented direct investigation of the effect of temperature on the diffuse band pattern".

Delayed Fracture of Materials under Tension, Torsion and Compression

BY C. GURNEY AND Z. BORYSOWSKI

Royal Aircraft Establishment, Farnborough

MS. received 1 December 1947; in final form 8 April 1948

ABSTRACT. Delayed fracture experiments were made in tension, torsion and compression, the times to fracture under a series of constant loads being determined. The materials tested were brass in ammonia vapour, polymerized methyl methacrylate (p.m.m.) in liquid carbon tetrachloride, and glass in air.

In all cases delayed fracture is attributed to the gradual spread of cracks caused by the preferential attack by the surrounding medium of the highly stressed material at the ends of the cracks. In tension, fractures were transverse and in torsion they were helical. In the case of brass and glass, subjected to tension and torsion, fracture of the respective materials occurred in not greatly differing times when the materials were subjected to the same maximum nominal principal tensile stresses. P.m.m. withstood higher nominal principal stress in torsion than in tension for equal times to fracture. The p.m.m. results were complicated by the absorption of carbon tetrachloride, which was greatly accelerated by tension or torsion stresses. During the time occupied by the tests, absorption was inappreciable when the p.m.m. was stress-free or stressed in compression.

In no case did delayed fracture occur in compression.

§ 1. INTRODUCTION

MANY materials if left under constant load will ultimately fracture. The delayed fracture is usually associated with attack of the substance under stress by the surrounding atmosphere. In the absence of stress, the attack may be inappreciable, but when the substance is stressed, very localized attack may take place, leading to the formation of cracks and ultimately to fracture. For metals, the phenomenon is usually called stress corrosion cracking. Under certain conditions of heat treatment, many aluminium- and magnesium-rich alloys suffer from stress corrosion cracking when exposed under stress to atmospheres containing chlorides; copper alloys are attacked by ammonia, and iron alloys by many salt solutions (Mears *et al.* 1944). The very remarkable delayed fracture of glass in air, for which a three-fold reduction in stress leads to a proportional increase in time to fracture of $10^6 : 1$, has been shown by Preston and his collaborators (1946) to be mainly due to the attack of the atmosphere. Organic plastics also give rise to delayed fracture, especially if surrounded by organic liquids or vapours. A good example is delayed fracture of polymerized methyl methacrylate (p.m.m.) when immersed in carbon tetrachloride, which gives inappreciable attack during the same period when the p.m.m. is stress free.

In nearly all the delayed fracture experiments so far made the materials have been tested in either tension or bending, whereas in practice materials have to withstand other stress systems. In the present experiments delayed fracture of materials subjected to tension, torsion and compression, has been investigated.

§ 2. MATERIAL

In the present series of experiments three materials were tested. The first was brass to Spec. B.S.251 of the following composition: Cu 62.4, Sn 1.2, Zn 36.4. Before machining, the material was heat-treated in air at 500°C. for 45 minutes and then cooled in air. After machining, the test pieces were heat-treated in air

at 270° C. for one hour and cooled in the furnace. Control tests were made in tension, torsion and compression. The nominal ultimate tensile stress when tested in air was 24.9 ton/in² and the 0.1% proof stresses were 9.5 tons/in² (tension) and 9.6 tons/in² (compression). The figures quoted are the means of four tests. The delayed fracture experiments were done in an atmosphere saturated with the vapour of aqueous ammonia solution of 35% w/w. NH₃.

The second material tested was polymerized methyl methacrylate (p.m.m.) sheet*. After machining, all test pieces were heated in air at 80° C. for half an hour and cooled in the oven. The ultimate tensile strength at a rate of increase of stress of about 5000 lb/in²/min. was about 10,000 lb/in². Delayed fracture experiments were made with the test pieces immersed in carbon tetrachloride.

The third material was a soda-lime glass in the form of $\frac{1}{4}$ in. diameter rods, the chief ingredients of which were present in the following percentages: silica 69, alumina 3, lime 6, magnesia 3, soda 18. It was annealed by the manufacturer, and subsequent measurements of internal stress by photoelastic methods indicated a mean value of surface compressive stress of about 600 lb/in² with a standard deviation of about 400 lb/in². The rods were packed without wrapping and tested without previous surface treatment. Delayed fracture experiments were made in air.

§ 3. APPARATUS

For delayed fracture experiments in tension and compression a spring-loaded cage was used. The test piece was screwed into plates at each end. Three screwed rods fitted with nuts and springs passed through each end plate. For the tension tests the springs were between the plates, one end of each bearing on an end plate and the other bearing on nuts screwed on to the rods. By tightening the nuts, the tension on the test piece was increased. For compression tests the springs were outside the end plates. In both tension and compression tests three Huggenberger tensometers were mounted on each test piece at angles of 120° to each other, and the nuts which loaded the springs were adjusted until the readings on each tensometer corresponded to the strain corresponding to the required stress. This arrangement ensured axial loading. For the torsion tests pulleys were screwed on to each end of the test pieces and located by lock nuts. Each pulley carried two separate cables attached to its rim at each end of a diameter. The cables were in contact with the pulley rims for an arc of about 90°, and then they left the pulley in opposite parallel directions. The directions of the cables on the pulleys at each end of a test piece were parallel and were arranged so that when all four cables were under equal tension the test piece was subjected to equilibrating end torques. The equality of the tensions was arranged by attaching the two upward cables to the opposite ends of a beam and by attaching the two downward cables to the opposite ends of another beam; the whole apparatus could be put inside the cages used for the tension and compression experiments, which were adapted to apply central loading to the two beams. For small loads the beams were pulled apart by dead loading.

The brass test pieces were 2 $\frac{3}{4}$ in. long threaded $\frac{1}{2}$ in. B.S.F. for a length of $\frac{3}{4}$ in. at each end, and reduced to 5.16 in. diameter over a central length of $\frac{5}{8}$ in. They had a central axial hole of 3.16 in. diameter running the full length of the test piece. The p.m.m. test pieces were 3 in. long threaded $\frac{1}{2}$ in. B.S.F. for a length

* Containing 8% dibutyl phthalate as plasticizer.

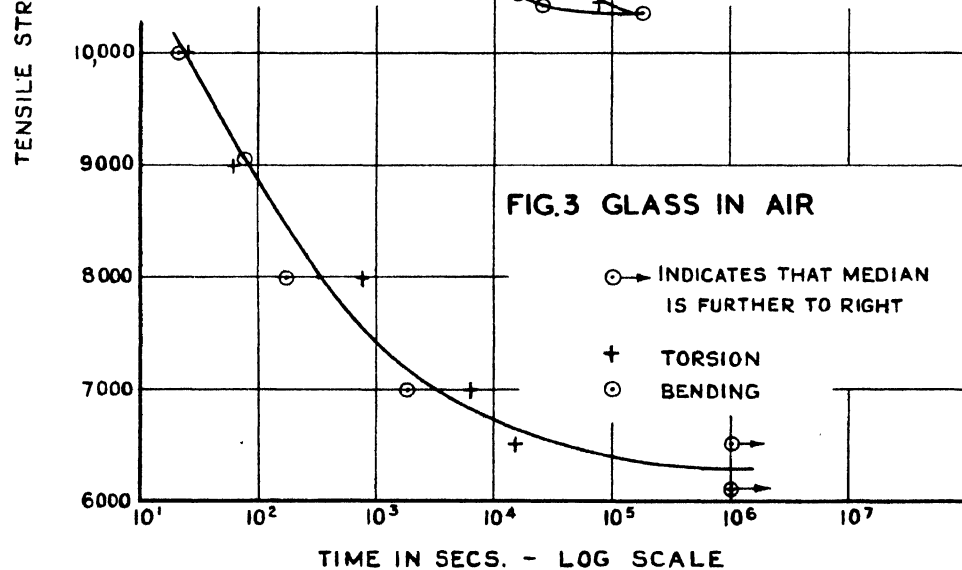
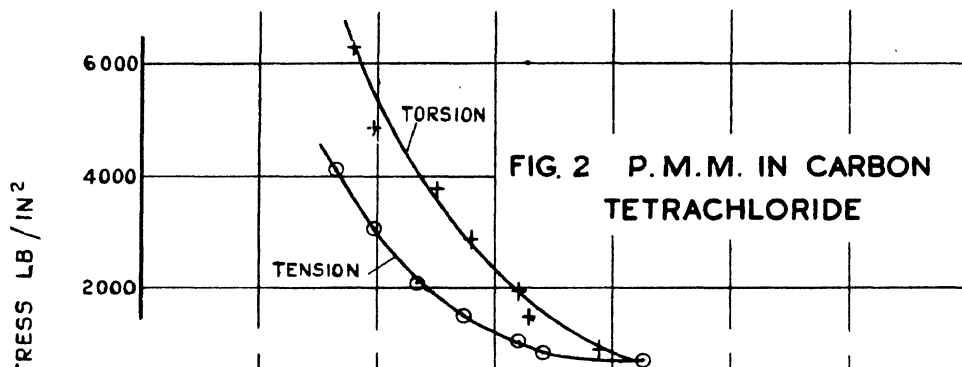
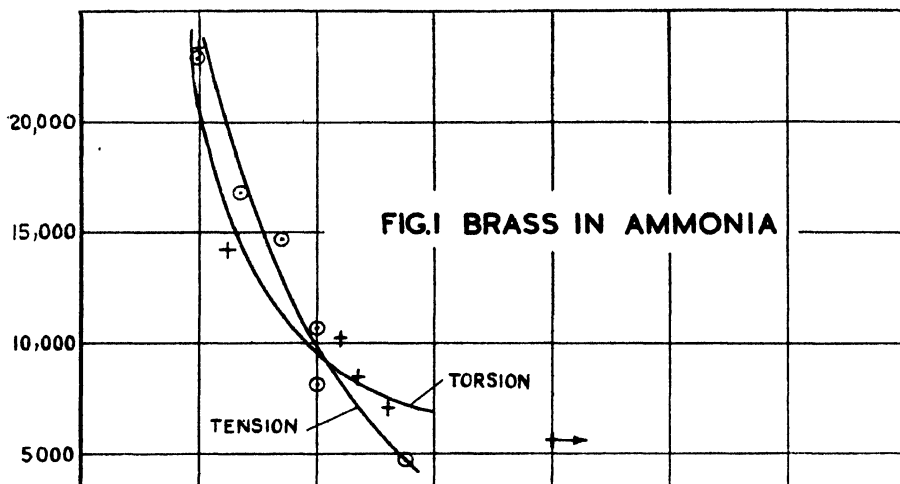
of 1 in. at the ends and reduced to $\frac{1}{4}$ in. diameter over a central length of $\frac{5}{8}$ in. The glass test pieces were round parallel rods of $\frac{1}{4}$ in. diameter. They were tested in bending instead of in tension, on account of the difficulty of obtaining axial loading. A four-point loading rig was used, the length subjected to uniform bending moment being 3 in. For torsion, the glass rods were cemented into 3 in. long steel sleeves with an organic cement. Pulleys were screwed on to the sleeves, and the tests were made in the same rig as that used for brass and p.m.m.

§ 4. EXPERIMENTS

The test pieces were set up in the apparatus already described and the times to fracture under a series of constant loads were recorded. The mean temperature during the tests was 22°C . This was thermostatically maintained to within $\pm 1^{\circ}\text{C}$. for the p.m.m. tests, but for the brass and glass tests the temperature was not so closely controlled. Figure 1 shows the results for brass, figure 2 those for p.m.m. and figure 3 those for glass. For the torsion tests the nominal maximum principal stress computed on the basis of a linear stress distribution is plotted against the logarithm of the time to fracture. For the tension tests the nominal average tensile stress has been plotted. There were five repetitions of each brass experiment, three repetitions of each p.m.m. experiment and twelve repetitions of each glass experiment. The points shown in figures 1, 2 and 3 are median points (half the test pieces failed at a lower stress). The median was chosen rather than the mean, as the former can be computed without waiting for all test pieces to break. The standard error of the median of a normal population is about 1.25 times that of the mean, so that the median is not a greatly worse statistic than the mean. The standard errors of the median points (on the log (time) scale) varied somewhat with stress, the standard error increasing with decreasing stress, and becoming very large at the lowest stresses where the curves tend to become parallel with the time axis. Except at the lowest stresses, the standard errors did not vary greatly with stress, and average standard errors of the median points have been computed from the mean of the ranges of results for each stress, omitting the results for the lowest stresses. The values for tension and torsion respectively were 0.09 and 0.08 for brass, 0.10 and 0.05 for p.m.m., and 0.5 and 0.5 for glass.

When brass in ammonia and p.m.m. in CCl_4 were stressed in compression at the highest stresses which could be applied without buckling the test pieces, no delayed fractures occurred and no cracks were noticed. Even when a short cylindrical piece of p.m.m. was immersed in CCl_4 and compressed so that it was deformed into a barrel shape, no cracking was observed. Compression tests on glass were not made, but it is known that glass is many times stronger in compression than in tension, so that delayed fracture in compression would not occur at stresses at which it occurred in tension.

When stress-free brass was left in the ammonia vapour it became covered with a blue deposit of more or less uniform thickness. When stress-free p.m.m. was left in CCl_4 no attack was detectable during the time occupied by the delayed fracture experiments, but after some months immersion at 22°C . the surface of the p.m.m. was seen to be swollen and covered with a gelatinous layer and the test piece had lost weight, showing that solution was taking place. The soda-lime glass was not noticeably attacked by the atmosphere when stressed or unstressed, but it is known that attack takes place after prolonged exposure. At elevated temperatures glass is rapidly attacked by steam.



§ 5. APPEARANCE OF FRACTURES

In the tension tests fracture of the brass test pieces started by the formation of approximately transverse intercrystalline cracks. Many test pieces had more than one crack of about the same depth, and final fracture often consisted in bending off the material between two cracks that had started from opposite sides of the test piece. There was no appreciable reduction in area. In the torsion tests, fracture of the brass test pieces started by the formation of cracks at about 45° to the axis at the test piece. Completion of the fracture usually involved bending and tearing of the test piece.

The tension fractures in the p.m.m. test pieces were transverse and the torsion fractures were helical. Examination of the fractured test pieces showed a very interesting phenomenon. Test pieces which had been subjected to comparatively low stresses showed fractures which appeared to start at the surface, as might be expected; but fracture in the test pieces which had been subjected to higher stresses started, not from the surface, but from a point in the interior. Apart from examination of the actual fractures, there was other unmistakable evidence that fracture started internally; after tests had been in progress for some time, both tension and torsion test pieces developed clearly visible internal cracks which were transverse in the case of tension and helical in the case of torsion.

The fractures of the glass rods call for no special comment as their appearance is the same as that of fractures occurring during quick-loading tests. The bending fractures were usually S-shaped and the torsion fractures were helical.

§ 6. DISCUSSION

In all three materials delayed fracture is regarded as being due to the gradual spread of cracks. The cracks were clearly visible in the brass and p.m.m. test pieces. They were not visible in the glass test pieces, but there are good reasons for attributing the delayed fracture of glass to the growth of very small cracks (Gurney 1947 a). When material containing cracks is stressed there is a concentration of strain energy at the ends of the cracks and consequently the material at the end of the cracks has a different chemical potential from that of the rest of the material; and under some stress systems it is preferentially attacked by any chemically active surrounding medium. Appropriate thermodynamic relations when the chemical attack takes the form of swelling and solution are given in a recent paper (Gurney 1947 b). Other cases could be calculated using similar methods. The thermodynamic relations are somewhat complex and there is no reason to suppose that any simple parameter of the stress system can be used to express results for different stress systems and different chemical substances. Nevertheless the fact that delayed fracture did not occur in compression suggests that the maximum principal tensile stress may prove useful in expressing results for simple stress systems such as unidirectional stress and torsion. Accordingly this parameter has been plotted as ordinate in figures 1, 2 and 3.

For brass and glass the curves for tension and torsion are not very different, and in the absence of better information the maximum principal tensile stress criterion might be used as a rough working rule. There are several factors which would tend to cause the tension and torsion curves for brass to differ. The assumption of linear stress distribution in torsion would be in error at the higher stresses. At the lower stresses cracks occupying a considerable part of the cross-section developed before fracture occurred. The nature of the stress distribution

would be entirely altered by the presence of these large cracks. In particular, the stress at the ends of the cracks, which presumably is an important factor affecting the rate of crack growth, might be very different in the torsion tests and the tension tests, even though the nominal maximum principal tensile stresses assuming uncracked material were equal. In the case of glass these factors are inoperative as the assumption of linear stress distribution in torsion is probably accurately fulfilled, and fracture occurs before cracks of appreciable length compared with the diameter of the test pieces have developed. The experimental points for glass in torsion and bending are not significantly different, and a single curve has been drawn to represent the results.

The difference between the tension and torsion curves for p.m.m. is greater than that between the corresponding curves for brass or glass. The factors which have been suggested to account for the difference in the case of brass are also operative for p.m.m., but it seems doubtful that these effects would be so much greater for the p.m.m. test pieces than for the brass test pieces as to account for the large difference. Examinations of the fractured p.m.m. test pieces showed the peculiarity already mentioned—that at high stresses fracture started internally. This is evidence for an additional factor affecting the p.m.m. results which finds a ready explanation. When stress-free p.m.m. is immersed in CCl_4 for prolonged periods, it is found that the surface is swollen by absorption of CCl_4 , and that some of the p.m.m. has actually gone into solution in the liquid. The first effect of immersing stress-free p.m.m. in CCl_4 therefore will be a gradual absorption of the CCl_4 accompanied by swelling, but after prolonged immersion solution in the CCl_4 takes place. The effect of stress of the effects can be deduced from general thermodynamic reasoning. Natural processes tend to a reduction of the free energy of systems, and part of the free energy of a system under load lies in the potential energy of the external forces. If a rod under tension absorbs fluid, the external forces do positive work, and therefore reduce their potential energy for two reasons. Firstly, there is a direct increase in the length of the rod due to swelling, and secondly, there is an increase in the length of the rod due to the reduction in Young's modulus consequent upon swelling. The effect of tension, therefore, is to increase the rate of swelling. The effect of torsion is also to increase the rate of swelling; the reduction in shear modulus of the swollen test pieces allows the end couples to rotate and do work.

The swelling of the surface material of a rod under tension or torsion causes two effects. The swollen material is weakened and therefore more likely to fracture; but its size is increased and its elastic moduli (E and G) are reduced, while the elastic constants and dimensions of the unswollen interior material remain the same. The swelling of surface material tends to set up a system of internal stress in which the surface material is in two-dimensional compression. Both in tension and torsion, therefore, the maximum principal tensile stress in the surface is reduced. A further reduction is caused by the change in elastic moduli. Absorption of liquid by surface layers therefore weakens them, but at the same time causes a reduction in the stress they have to withstand.

These ideas lead to a plausible explanation of the experimental results that at low stresses fracture of the p.m.m. test pieces started from the outside, and at high stresses fracture started from the inside. It is only necessary to assume that at low stresses the weakening of surface material by absorption of CCl_4 exceeds the reduction in the stress associated with the absorption; fracture would then start

from the outside. On the other hand, at high stresses the reduction in stress due to swelling and change in elastic moduli exceeded the weakening of the material due to absorption, so that fracture started in the interior, where the stress was already increased by the internal tension stress caused by the swelling of the surface layers.

The broad conclusions to be drawn from the experiments are that delayed fracture does not occur in compression. For brass in ammonia and glass in air, delayed fracture in tension and torsion occurs in approximately equal times at equal principal tensile stresses. P.m.m. in carbon tetrachloride will withstand higher principal tensile stresses in torsion than in tension for equal times to fracture.

ACKNOWLEDGMENT

The delayed fracture experiments on glass were made by Mr. S. Pearson. Measurement of internal stress was done at the Research Laboratories of the General Electric Company.

REFERENCES

- GURNEY, C., 1947 a, *Proc. Phys. Soc.*, **59**, 169 ; 1947 b, *Ibid.*, **59**, 629.
 MEARS, R. B., BROWN, R. H., and DIX, E. H., 1944, *Symposium on stress corrosion cracking of metals* (Philadelphia : A.S.T.M.).
 PRESTON, F. W., BAKER, T. C., and GLATSHART, J. L., 1946, *J. Appl. Phys.*, **17**, 162.

β -Disintegration and the Sargent Diagram

BY N. FEATHER AND H. O. W. RICHARDSON

Department of Natural Philosophy, University of Edinburgh

MS. received 6 May 1948; to be read 5 November 1948

ABSTRACT. Sargent diagrams constructed for negative-electron-active species having $80 < Z < 94$ (as also for the restricted group with $80 < Z < 84$), and for both negative-electron-active and positron-active species having $Z < 22$, are discussed in the light of theory. The majority of first-forbidden transitions in the group of heaviest radioelements give a Sargent line parallel to the allowed line, corresponding to disintegration constants smaller by a factor of roughly 10 only than the disintegration constants of allowed transitions of equal energy. Transitions with no parity change between states of zero spin appear to give a second-forbidden line also parallel to the allowed line.

With the lightest radioelements "super-allowed" transitions are identified which require Wigner selection rules for their interpretation, but there is also a group of transitions which appears to be ordinarily allowed (i.e. subject to spin and parity selection rules only). Compared with these allowed transitions in the lightest elements, allowed transitions in the heaviest radioelements are less probable by a factor of about three only, when comparison is made for equal energy release and account is taken of the effect of the coulomb field of the nucleus.

On the assumption that all ^(even)nuclei have even parity and zero spin in the ground state it is concluded that the β -particles of maximum energy observed with RaB, RaC and MsTh₁ are emitted in transitions to excited states of the final nuclei RaC, RaC' and MsTh₂, respectively. This conclusion makes it likely that the accepted value of the total γ -ray emission from RaB (Gray 1937) is too small, that the total disintegration energy Ra.C.C' is 3.78 mev. and that MsTh₁ emits a low-energy γ -radiation which has not as yet been detected. On the same assumption it is concluded that the disintegrations UX₁, UX₂ and UX₂. U_{II} possess features at present unknown or insufficiently understood.

A REVISED Sargent diagram for the heavy radioelements has recently been published by one of us (Feather 1948) showing strong evidence for a hitherto unidentified "line", almost parallel to the line for allowed transitions, and not very distant from that line. If the suggestion that the new line can belong only to first-forbidden transitions be accepted, then obviously the degree of forbiddenness corresponding to the other lines, previously identified, and lying lower on the diagram, must be revised—the first-forbidden line of the old diagram becoming the second-forbidden line of the new diagram, and so on. It is one object of this note to point out that this assignment has the support of theory: the occurrence of first-forbidden transitions, differing relatively little in probability from allowed transitions of the same energy, has been predicted by Konopinski and Uhlenbeck (1941), at least for the heavy elements. For the lightest radioelements, according to these authors, such transitions should not occur: the calculations show that a much larger difference in transition probability is to be expected as between allowed and first-forbidden disintegrations of equal energy.

The formal predictions of the theory are briefly as follows. Let R_{01} be the ratio of the disintegration constants, $\lambda_0(W)$, $\lambda_1(W)$, for allowed and first-forbidden transitions in which the total disintegration energy (including the rest-energy of the electron) is Wmc^2 . Let Z be the charge number of the β -active nucleus, and let the nuclear radius be $\rho\hbar/2\pi mc$. Then when

$$\alpha Z \gg 2\rho W, \quad \dots\dots(1)$$

α being the fine-structure constant, $1/137$, we have, alternatively,

$$R_{01} \sim (\alpha Z/2)^{-2} \quad \dots\dots(2)$$

or

$$R_{01} \sim (\rho W)^{-2} \quad \dots\dots(3)$$

depending upon the change of spin in the transition. If I is the total spin quantum number of the nucleus, (2) applies to those transitions which involve a change of parity of the nuclear state and for which $\Delta I = 0$ or ± 1 and (3) applies when $\Delta I = \pm 2$. (Here it might be noted that first-forbidden transitions with $\Delta I = \pm 2$ are possible only on the basis of the axial vector and tensor forms of interaction, which are governed by Gamow-Teller selection rules.) The theory also predicts that the shape of the energy spectrum should approximate to the allowed shape when (2) applies, and differ considerably from this shape, showing a deficit of medium-energy β -particles and an excess of high-energy particles, when (3) is valid.

For the heavy elements ($80 < Z < 94$) for which (1) is amply fulfilled, (2) points to a first-forbidden line parallel to the line for allowed transitions and distant, approximately, by $\Delta \log \lambda = 1.0$ ($8.5 < (\alpha Z/2)^{-2} < 11.7$)—and (3) to a lower-lying line, quite steeply inclined to the allowed line at its upper end, but still representing transitions which are only once forbidden. On this analysis of the situation there is clearly a good case for our identifying the newly-recognized line with the first-forbidden line predicted by (2): it exhibits the close parallelism with the experimental line for allowed transitions which is required, and the absolute value of the mean separation ($0.75 < \Delta \log \lambda < 1.1$) agrees remarkably well with prediction.

Having elicited this measure of agreement, the next object of this note is to continue the comparison of experiment and theory in respect of transitions of higher degrees of forbiddenness.

(Marshak 1942, Greuling 1942), who find in it evidence for narrowing the choice of possible interaction forms to the polar vector (Fermi rules) or tensor (Gamow-Teller rules) forms—but we attempt to follow the assumption that a closely similar value of a is appropriate to the other second-forbidden transitions to which (4) applies (within our chosen range of Z). Curves IIa and IIa', therefore, have been constructed, the latter according to (6) through the representative point for RaE, and the former parallel to it and higher in the diagram by $\Delta \log \lambda = 0.699$ ($= \log 5$). In constructing these curves, which are of course drawn with respect to the allowed curve O as reference curve, the small variations in Z and ρ , possible for the various points in the diagram, have been neglected. Curve O, it should be pointed out, is appropriate only when $\Delta I = 0$ or $+1$; curve O' belongs to those allowed transitions for which $\Delta I = -1$ ($\Delta \log \lambda = 0.477$). Similarly for the curves Ia and Ia' which belong to the first-forbidden transitions governed by (2): Ia is applicable when $\Delta I = 0, +1$ and Ia' when $\Delta I = -1$ *.

In considering the whole diagram at this stage it will be seen that one point only lies between curves Ia' and IIa, and that not more than two "observed" points lie below curve IIa'. Formally, therefore, at least 19 of the 22 "observed" points would appear to be satisfactorily accounted for—even though four of the five "unobserved" points, which lie below curve IIa', require further consideration. We shall return later to the general question of the "unexplained" points, for the present we direct attention to the group of four transitions which has emerged as second-forbidden according to (4), through our construction of curves IIa and IIa' as above described. In addition to the ground-to-ground state transition RaE.F, we have the "unobserved" ground-to-ground state transition RaD.E (for which a disintegration energy of 60 ± 4 kev. and an intensity of < 0.01 has been assumed) and two transitions ThC.C' (those leading to the ground state of ThC' and to the excited state of 726 kev. energy, respectively).

In the first place it is extremely satisfactory that the two ground-to-ground state transitions belonging to the successive disintegrations $\text{RaD} \rightarrow \text{RaE} \rightarrow \text{RaF}$ should be found grouped together in this way. The initial and final nuclei, $^{210}_{82}\text{RaD}$ and $^{210}_{84}\text{RaF}$, are both $\begin{smallmatrix} \text{even} \\ \text{even} \end{smallmatrix}$ nuclei, and in the ground state each would be expected to have $I=0$ and even parity. Thus the β -disintegrations RaD.E, RaE.F which involve the same intermediate state (the ground state of RaE) must be expected to be characterized by equal and opposite values of ΔI and by the same parity-change (either each "yes" or each "no"). The two transitions should, in short, be forbidden to identical degree. This is exactly what is found. (To be precise, the point for the ground-to-ground state transition RaD.E ought to lie on curve IIa, if that for RaE.F is on curve IIa', but it cannot be held that the distinction is important having regard to the uncertainty as regards intensity and the extent of the extrapolation involved in the comparison just made.)

We next discuss the two points for ThC.C'. In the successive disintegrations $\text{ThB} \rightarrow \text{ThC} \rightarrow \text{ThC}'$ the situation in respect of A and Z is similar in all respects to that just considered: $^{212}_{82}\text{ThB}$ and $^{212}_{84}\text{ThC}'$ are each of them $\begin{smallmatrix} \text{even} \\ \text{even} \end{smallmatrix}$ nuclei †.

* According to Marshak (1942) curves O' and Ia' belong strictly to those transitions characterized by $I=1 \rightarrow I=0$. Other transitions for which $\Delta I = -1$ should have representative points between curves O and O', if allowed, or, if first-forbidden, between Ia and Ia'. See letter by Strachan in this issue, p. 431.

† The classification $\begin{smallmatrix} \text{even} \\ \text{even} \end{smallmatrix}$, $\begin{smallmatrix} \text{even} \\ \text{odd} \end{smallmatrix}$, etc., of this paper refers throughout to neutron and proton numbers, i.e. is a classification in respect of $\begin{pmatrix} A \\ Z \end{pmatrix}$, not of $\begin{pmatrix} A \\ N \end{pmatrix}$.

thus we should expect the ground-to-ground state transitions ThB.C, ThC.C' to be identically forbidden. We have mentioned that the point for ThC.C' falls within the limits of curves IIa and IIa', but in this case the corresponding point for the previous disintegration does not. In fact the point for the ground-to-ground state transition ThB.C is the single point which on figure 1 lies mid-way between curves Ia' and IIa. Formally this point might be interpreted as being on a curve Ib, determined by (3)—then $|\Delta I|$ would have the same value, 2, for the transitions which are thus associated—but we should not in this way solve the problem, for the parity changes would then be different. If our assumptions regarding spin and parity of the ground states of ^(even) nuclei are correct, there would, in fact, appear to be only one solution—the ground-to-ground state transitions ThB.C, ThC.C' must be of a special type which we have not hitherto considered. This suggestion might be thought to be strengthened by the empirical fact that the two representative points concerned are very closely the same distance below the line for allowed transitions—and, from the point of view of theory, by the fact that transitions $I=0 \rightarrow I=0$ constitute a naturally special class when certain of the five standard interaction forms are employed (this is untrue only with the scalar and polar vector interactions, which carry the Fermi rules). We adopt this suggestion, then, drawing the line OO₂ parallel to O and distant by $\Delta \log \lambda = 1.85$ —and we make the further natural assumption that the new line (OO₂) is characteristic of a whole class of $I=0 \rightarrow 0$ transitions generally (within the chosen range of Z). That this class of $0 \rightarrow 0$ transitions is that for which there is no change of parity may be deduced from the following considerations. The representative point for the main ThB.C transition falls on the allowed line. Thus there is no parity change in this β -transition, which leads to the 238 kev. state of ThC. The γ -ray emitted in transition from this state to the ground state is converted internally as a magnetic dipole radiation. The excited and ground states, therefore, have like parity. The ground-to-ground state β -transition ThB.C is likewise, therefore, a transition without change of parity.

Now $0 \rightarrow 0$ transitions without change of parity are permitted according to the scalar and polar vector forms of interaction as ordinary allowed transitions, and such transitions are completely forbidden according to the axial vector and pseudoscalar forms of interaction. On the tensor form only are they permitted merely as (second) forbidden transitions. It appears therefore that our line OO₂ can only be explained, on the basis of a single (unmixed) interaction, as a "second-forbidden" line for $0 \rightarrow 0$ transitions obeying Gamow-Teller selection rules and describable in detail using the tensor form of interaction. It is satisfactory that theory in this case predicts a line which is parallel to the allowed line (Konopinski 1943). It is also satisfactory that the distance of OO₂ from the allowed line is approximately twice the distance of Ia from this line, for amongst the transitions to which Ia might belong are the $0 \rightarrow 0$ transitions with change of parity which, on the basis of the tensor form of interaction, behave as first-forbidden. Using an obvious notation we might in fact say that a line OO₁ coincides roughly with the line labelled Ia in figure 1.

We have obviously digressed from our discussion of the two points for ThC.C' ; returning now to this specific topic we may note that whilst one of them has been taken out of the group to which the curves IIa and IIa' apply, there is nothing to contradict the identification of the other as the representative point of a second-forbidden β -disintegration having $|\Delta I| = 2$.

A third case of successive β -disintegrations for which the initial nucleus is an (even) nucleus is provided by the sequence $\text{RaB} \rightarrow \text{RaC} \rightarrow \text{RaC}'$. Although it cannot be claimed that sufficient is known regarding these much-studied disintegrations to provide full information for a discussion, if we make the same assumptions as before regarding spin and parity certain conclusions follow without question. The most important of these is that the ground-to-ground state transition RaB.C is "unobserved". This is in line with the opinion originally held (cf. Ellis 1934), but it has been rendered doubtful by results of Gray (1937). The present argument is as follows. We may assume for the moment that it is not certain whether the ground-to-ground state transition $\text{RaC.C}'$ is that of 3.173 mev. maximum energy given by the experimentally determined end-point of the RaC β -particle spectrum, or an unobserved transition of 0.606 mev. higher energy (Ellis 1934, Bothe and Maier-Leibnitz 1937), but in either case it appears that the transition is at least second-forbidden. On the other hand, the experimentally determined end-point of the RaB β -particle spectrum at 0.72 mev. energy (Constantinov and Latyshev 1941) belongs to a transition which is obviously allowed. This is enough to show that the two transitions are not consecutive; they cannot have the ground state of RaC as a common intermediate state; in other words the ground-to-ground state transition RaB.C is "unobserved".

Concerning the disintegration $\text{RaC.C}'$ again, it has just been stated that it is not certain, on the direct experimental evidence, whether or not the ground-to-ground state transition is observed. We can now use figure 1 to establish the conclusion that it is almost certainly not observed. Evidence from the long-range α -particles of RaC'.D shows that a state of RaC' of 2.88 mev. energy is excited in the previous β -disintegration, and it is practically certain that de-excitation of this state involves the emission of the strong (quadrupole) γ -rays of 1.12 and 1.76 mev. energy in cascade (Rutherford, Lewis and Bowden 1933). The absolute intensities of these γ -radiations have been shown to be closely the same (Latyshev 1947) and calculated values are such that an initial excitation of the 2.88 mev. state of at least 20% must be assumed. If the total energy of disintegration is 3.17 mev., excitation of the state in question involves a β -particle partial spectrum of 0.29 mev. maximum energy only, and the highest excitation possible (assuming an allowed transition) is then about 1.5%. On the other hand, a partial spectrum of (0.29 + 0.606) mev. maximum energy would in fact give 22.5% or 67.5% excitation, depending upon whether curve O' or O were appropriate to the case. Curve O' would apply if the transition were one for which $\Delta I = -1$; on this matter all that can be said is that it is very unlikely that ΔI is positive, though it might be zero. In any case, however, the indication is clear enough; the total disintegration energy must be considerably greater than 3.17 mev.; thus the ground-to-ground state transition $\text{RaC.C}'$ is not observed.

It is worth while discussing the unobserved transitions in this disintegration somewhat further—for there is clear evidence that there are many of them. Thus Rutherford, Lewis and Bowden (1933) detected twelve groups of long-range α -particles from RaC' —in this way providing evidence for the excitation, direct or indirect, of twelve states other than the ground state in the disintegration $\text{RaC.C}'$ —and the same authors were forced to postulate the excitation of at least four other states in order to account even for the stronger lines in the γ -ray spectrum of RaC . Now, we have already suggested somewhat more than 20% excitation for the state of highest energy (2.88 mev.), and the general shape of the β -particle spectrum (Constantinov and Latyshev 1941) is inconsistent with the direct

excitation of more than a few other states to any comparable extent (say about 25% for the state of 0.606 mev., and roughly 50% for a state or states at about 2.2 mev. excess energy). So our statement may be made more precise: probably not more than four of the sixteen excited states which are regarded as definitely established are excited by direct β -transition $\text{RaC.C}'$ to more than 10%. Several of these states cannot be excited to more than a few tenths per cent at the most. Assuming that in the ground state the nucleus $^{214}_{84}\text{RaC}'$ has $I=0$ and even parity, we thus conclude that there are probably other states, of higher energy than that of 1.41 mev., which like the latter state (Fowler 1930) also have $I=0$. Some of these may not be involved in γ -ray emission to any appreciable extent. Depending upon whether $\Delta I=0$ or -1 for the transition exciting the 2.88 mev. state, bearing in mind the nature of the cascade γ -ray emission from this state, we are led to assign $I=4$ or $I=5$ to the ground state of RaC . Thus these unobserved transitions must be at least third-forbidden ($|\Delta I| \geq 4$).

The general features of the level scheme begin to emerge from this discussion, but clearly it would be unprofitable to continue it further until the results of experimental studies of β - γ and γ - γ correlation provide more detailed information. Here it should merely be remarked that only three points are plotted for RaC on figure 1 ($E=3.173$ mev., $23 \pm 5\%$; $E=1.51 \pm 0.02$ mev., $40 \pm 10\%$; $E=3.78$ mev., $<0.5\%$): to have included all the "unobserved" points would have been neither practicable nor very helpful at this stage.

It has already been stated that not more than two of the "observed" points of figure 1 lie below the line IIa' . One of these is the point for the 3.173 mev. transition $\text{RaC.C}'$ which has just been discussed at length, the other is the point for the ground-to-ground state transition $^{204}_{81}\text{Tl} \xrightarrow{3.5\gamma} ^{204}_{82}\text{Pb}$. This point almost certainly belongs either to curve IIb' or to curve IIIa' —the accents reflecting the fact that since the final nucleus is an $^{(\text{even})}_{82}\text{Pb}$ nucleus ΔI is presumably negative in this case. On the basis of either attribution $\Delta I = -3$ (parity being unchanged if IIb' applies and changing if IIIa' is in question). Again, since the energy dependence of λ is the same according to either assumption (compare (5) with $R_{03} \sim (\alpha Z/2)^{-2}(\rho W)^{-4}$ which is applicable to curve IIIa') we may, without further decision between the possibilities, construct a curve through the point under consideration and another curve representing values of λ greater by a factor of 7 ($2|\Delta I| + 1$, with $|\Delta I| = 3$). These curves are shown on figure 1 and appropriately labelled. It will be noted that they intersect curve IIa' in the general region of the point for RaC ($E=3.173$ mev.), thus no definite assignment of this point appears possible on the basis of the diagram alone. A similar remark is obviously true of the unobserved point representing the 2.38 mev. transition ThC'' . D. As regards the "unobserved" points for the ground-to-ground state transitions RaC'' . D, ThC'' . D and $\text{RaC.C}'$ (see above), it would appear from their positions on the diagram very likely that these belong to transitions which are at least fourth-forbidden (see Richardson 1948).

Figure 1, it will be remembered, was constructed by omitting the 23 points representing β -transitions in elements having $87 \leq Z \leq 93$ from the Sargent diagram recently published. We may reconsider these transitions briefly now, making the following general observations:

- (1) 7 of the omitted points would lie, on figure 1, either between curves O and O' or somewhat above curve O.

(2) 8 of them would lie between curves Ia and Ia' (thus it happens that the evidence that these curves are significant is statistically stronger on the basis of the higher- Z group data than on the lower- Z group data).

(3) 1 point (belonging to a rare mode of the disintegration $UX_2 \cdot U_{II}$) would appear to lie on the line OO_2 .

(4) 4 points would lie between curves IIa and IIa'.

(5) 1 point would be found somewhat below the line $IIb' - IIIa'$ (this is the only point plotted for $MsTh_2$, the disintegration of which is obviously almost as complex as that of RaC but much less thoroughly studied).

(6) The remaining two ("unobserved") points would appear to belong to transitions which are third- or fourth-forbidden (these are the points for the ground-to-ground state transitions $UZ \cdot U_{II}$ and ${}^{238}_{93}Np \cdot {}^{238}_{94}Pu$).

Clearly there is nothing in these general observations to suggest that our detailed treatment of the points included in figure 1 would fail if applied to the omitted points; the effect of the change in average Z value between the two groups (from $Z = 82$ to $Z = 91$) in fact appears to be even smaller than might have been predicted.

But there is one particular case which merits further discussion, for on the evidence at present available it seems that a real difficulty is involved. The disintegration sequence $UX_1 \rightarrow UX_2 \rightarrow U_{II}$ between the $({}^{even}_{even})$ nuclei ${}^{234}_{90}UX_1$ and ${}^{234}_{92}U_{II}$ provides the case in point. These disintegrations have been very carefully studied by Bradt and others (Bradt, Heine and Scherrer 1943, Bradt and Scherrer 1945, 1946), and, in spite of the complications due to the $UX_2 - UZ$ isomerism, experimentally it would seem that the evidence is definite and in the main complete. But the positions of the representative points for the partial spectra of highest energy in the disintegrations $UX_1 \cdot UX_2$, $UX_2 \cdot U_{II}$ certainly lie on different Sargent curves on the revised diagram—the former between curves Ia and Ia' and the latter on curve O. There is no evidence at present against the obvious assumption that the partial spectra in question characterize truly consecutive transitions, or against the assumption that the nucleus U_{II} is left in the ground state after the second of these transitions, yet the Sargent diagram analysis points to a change of parity in the first transition ($UX_1 \cdot UX_2$) and no change of parity in the second ($UX_2 \cdot U_{II}$). The apparent overall change of parity as between the initial and final $({}^{even}_{even})$ nuclei disclosed by this analysis is the difficulty already mentioned.

There is another more subtle difficulty which concerns the disintegration $UX_2 \cdot U_{II}$ considered in isolation from the preceding transformation. If the two component radiations of about 0.95 mev. and 1.5 mev. energy identified by Bradt and Scherrer (1945, 1946) were the only γ -radiations emitted by UX_2 the difficulty would not arise: observed intensities fit well with the assumption that for these radiations excitation of the γ -emitting states (of U_{II}) is by first-forbidden transitions, "excitation" of the ground state being by an allowed transition as above described. But there are two or three other radiations (of 782, 822 and possibly 806 kev. energy, respectively) to be taken count of, and there is strong circumstantial evidence that some or all of these radiations are emitted in $0 \rightarrow 0$ transitions leading to the ground state. The serious difficulty in this interpretation is that, if the emitting states and the ground state have the

same spin ($I=0$) and parity, excitation of the emitting states by β -disintegration should be to the extent of about 20% whereas in fact it is almost certainly less than 0.5%. Even the assumption of different parities would not entirely remove the discrepancy. Empirically the representative points for the β -transitions exciting the emitting states in question (only one such point was plotted in the Sargent diagram recently published) appear to lie on the curve OO_2 ($0 \rightarrow 0$ transition, no parity change), as mentioned in our general observation (3) above. This, of course, is consistent with the assumption that the subsequent radiations involve $0 \rightarrow 0$ transitions to the ground state of the $(\text{even})_{\text{even}}$ nucleus ${}^{234}_{92}\text{U}_{\text{II}}$: it is rather the main β -transition $\text{UX}_2 \cdot \text{U}_{\text{II}}$ which does not appear to fit into the general interpretative scheme—it cannot similarly be described as a $0 \rightarrow 0$ transition between states of even parity in view of the position of the representative point on the allowed curve O of the revised diagram. It would seem, therefore, that unless our general assumption regarding spin and parity of $(\text{even})_{\text{even}}$ nuclei in the ground state is at fault, some new feature of the $\text{UX}_1 \cdot \text{UX}_2$, $\text{UX}_2 \cdot \text{U}_{\text{II}}$ disintegrations remains to be discovered. More detailed studies of the known γ -radiations in emission, and a search for a possibly highly converted low-energy γ -radiation of high intensity, would appear to be called for. It might be mentioned here that no difficulty arises in assigning degrees of forbiddenness to the various β -transitions in the associated disintegration $\text{UZ} \cdot \text{U}_{\text{II}}$: energies and intensities of excitation, and the nature of the γ -radiations, can be satisfactorily explained on the assumption that the β -transition of energy 0.45 mev. is allowed, that of energy 1.2 mev. second-forbidden and the unobserved ground-to-ground state transition of about 1.95 mev. energy third- or, more probably, fourth-forbidden.

There remains finally the case of the successive disintegrations $\text{MsTh}_1 \rightarrow \text{MsTh}_2 \rightarrow \text{RdTh}$, the last case amongst the “classical” radioelements of successive β -disintegrations between initial and final nuclei which are both $(\text{even})_{\text{even}}$ nuclei. Here, as already mentioned, information is far from complete, but it is complete enough to sustain one clear prediction, so long as our basic assumptions remain valid. According to Lee and Libby (1939) the maximum energy of the β -particles of MsTh_1 is 54 ± 4 kev. On the basis of a single disintegration mode this implies a first-forbidden transition according to our revised diagram. On the other hand, the ground-to-ground state transition $\text{MsTh}_2 \cdot \text{RdTh}$ is at least second and probably more highly forbidden (see our general observation (5) above). Thus we must anticipate that the ground-to-ground state transition $\text{MsTh}_1 \cdot \text{MsTh}_2$ is unobserved, that is, that there is an intense (soft) γ -radiation from MsTh_1 still to be discovered.

Our discussion of the Sargent diagram for the heaviest radioelements has been based on the tacit assumption that the only selection rules involved are those concerned with the parity and total angular momentum of the nucleus—and empirically there is no indication that these rules are inadequate. With the lightest elements, however, there is abundant evidence that further rules are required, and Wigner (1939) and others have formulated such additional rules employing a “super-multiplet” classification of nuclear energy states. The empirical evidence for the necessity of some such extension of theory is obvious from figures 2 and 3. These diagrams refer, respectively, to electron-active and positron-active species having $Z < 22$ *. In the absence of additional selection

* No point is plotted on figure 2 for ${}^8\text{Li}$, the β -spectrum of which cannot be resolved into a finite number of discrete partial spectra because the disintegration energy varies continuously from zero to about 12 mev. For a full discussion see Wheeler (1941).

rules we should expect the allowed curves to be well-defined, and that they should occupy identical positions on the two diagrams, the effect of the coulomb field of the nucleus being all but inappreciable for elements in this range of Z .

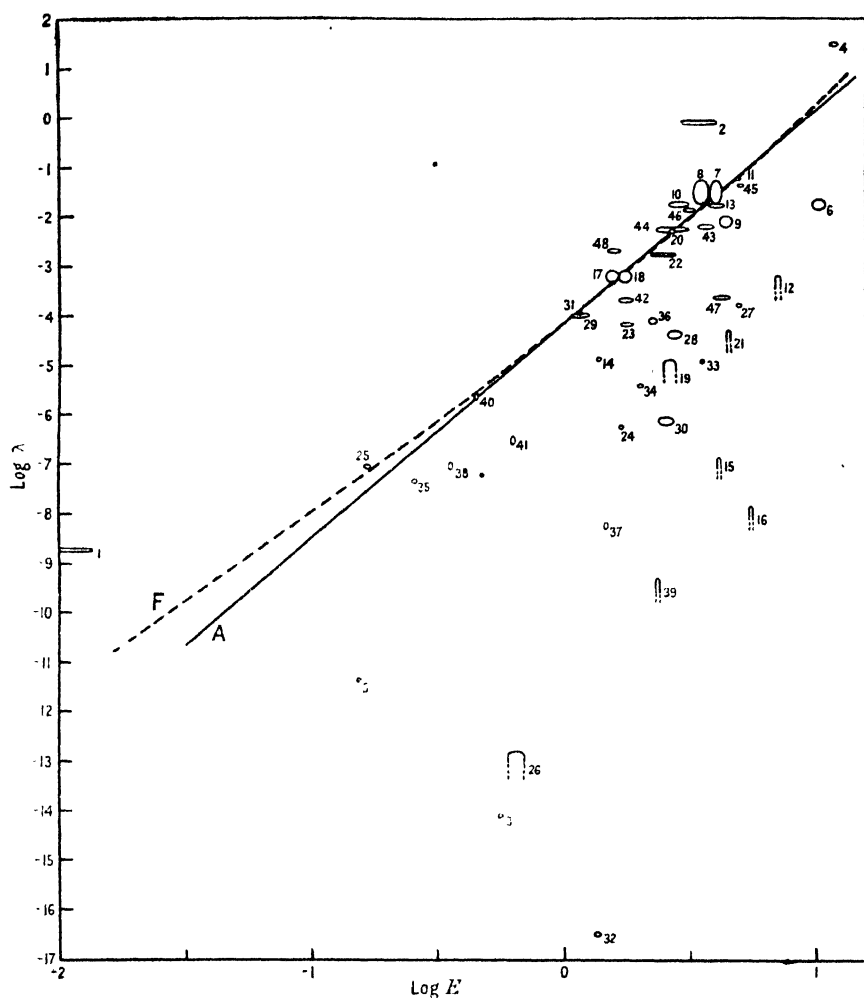


Figure 2. λ in sec^{-1} ; E in mev.

Key for figure 2

Species	Points	Species	Points	Species	Points	Species	Points
^3_1H	1	$^{23}_{10}\text{Ne}$	13	$^{34}_{15}\text{P}$	45, 46	$^{45}_{20}\text{Ca}$	35
^6_2He	2	$^{24}_{11}\text{Na}$	14, 15, 16	$^{35}_{16}\text{S}$	25	$^{46}_{20}\text{Ca}$	36
$^{10}_4\text{Be}$	3	$^{25}_{11}\text{Na}$	43, 44	$^{37}_{16}\text{S}$	47, 48	$^{46}_{21}\text{Sc}$	37, 38, 39
$^{12}_5\text{B}$	4	$^{27}_{12}\text{Mg}$	17, 18, 19	$^{36}_{17}\text{Cl}$	26	$^{47}_{21}\text{Sc}$	40
$^{14}_6\text{C}$	5	$^{28}_{13}\text{Al}$	20, 21	$^{38}_{17}\text{Cl}$	27, 28, 29	$^{48}_{21}\text{Sc}$	41
$^{16}_7\text{N}$	6, 7, 8	$^{29}_{13}\text{Al}$	22	$^{41}_{18}\text{A}$	30, 31	$^{40}_{21}\text{Sc}$	42
$^{18}_8\text{O}$	9, 10	$^{31}_{14}\text{Si}$	23	$^{40}_{19}\text{K}$	32		
$^{20}_9\text{F}$	11, 12	$^{32}_{15}\text{P}$	24	$^{42}_{19}\text{K}$	33, 34		

Comparison of the diagrams shows, however, that these predictions are far from the truth. There is the appearance of an allowed line (A) in figure 2—if we leave on one side for a moment the points for the β -active species ^3H , ^6He and ^{12}B —though it is not so well defined as the allowed line for the heavy radioelements in figure 1, but if this line is transferred from figure 2 to figure 3 (A') it is seen to fall well below the majority of points on that diagram. By contrast, in figure 3

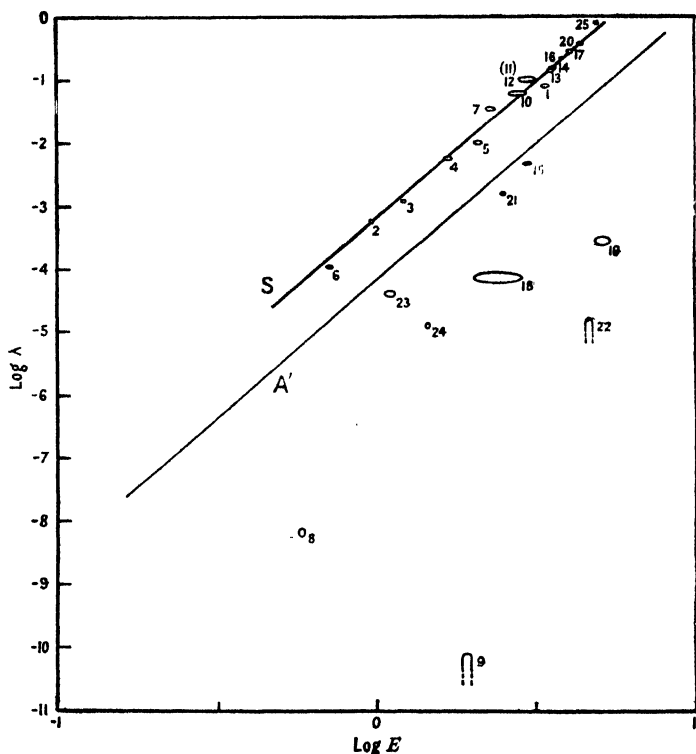


Figure 3. λ in sec^{-1} ; E in Mev.

Key for figure 3.

Species	Points	Species	Points	Species	Points	Species	Points
$^{10}_6\text{C}$	1	$^{19}_{10}\text{Ne}$	7	$^{29}_{15}\text{P}$	14	$^{38}_{19}\text{K}$	21, 22
$^{11}_6\text{C}$	2	$^{22}_{11}\text{Na}$	8, 9	$^{30}_{15}\text{P}$	15	$^{41}_{21}\text{Sc}$	25
$^{13}_7\text{N}$	3	$^{23}_{12}\text{Mg}$	10	$^{31}_{16}\text{S}$	16	$^{43}_{21}\text{Sc}$	23
$^{15}_8\text{O}$	4	$^{25}_{13}\text{Al}$	(11)	$^{33}_{17}\text{Cl}$	17	$^{44}_{21}\text{Sc}$	24
$^{17}_9\text{F}$	5	$^{26}_{13}\text{Al}$	12	$^{34}_{17}\text{Cl}$	18, 19		
$^{18}_9\text{F}$	6	$^{27}_{14}\text{Si}$	13	$^{35}_{18}\text{A}$	20		

there is a perfectly well-defined line (S) parallel to A', which line we can only classify empirically as "super-allowed". In this connection the treatment of Wigner is obviously significant: of the 15 points defining our super-allowed line, 12 belong to the ground-to-ground state transitions of 12 of the 16 ($^{2Z}_{Z-1}$) species $^{11}_6\text{C}$ to $^{41}_{21}\text{Sc}$ (and there is fairly clear indication from the known periods that the corresponding points for three at least of the remaining four species of this

sequence, viz. $^{21}_{11}\text{Na}$, $^{25}_{13}\text{Al}$ and $^{30}_{20}\text{Ca}$ —only $^{37}_{19}\text{K}$ is at present unknown—also fall on this line)—and the other three represent the ground-to-ground state transitions of $^{10}_6\text{C}$, $^{18}_9\text{F}$ and $^{26}_{13}\text{Al}$. According to Wigner, only transitions belonging to the classes

$$\begin{pmatrix} 2Z \\ Z-1 \end{pmatrix} \xrightleftharpoons[\beta-]{\beta+} \begin{pmatrix} 2Z \\ Z-1 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 2Z \\ Z+1 \end{pmatrix} \xrightleftharpoons[\beta-]{\beta+} \begin{pmatrix} 2Z \\ Z \end{pmatrix} \xrightleftharpoons[\beta-]{\beta+} \begin{pmatrix} 2Z \\ Z-1 \end{pmatrix}$$

with Z odd are not necessarily forbidden by any rules. It would appear that for $Z < 22$ * all positron-active species falling into the first of these classes undergo super-allowed transitions (from ground to ground state) and that some of the species of the second class behave similarly, though others do not. No super-allowed transitions, however, occur for positron-active species which belong to neither class. Amongst the partially forbidden transitions of species of the second class we might mention the ground-to-ground transitions of $^{22}_{11}\text{Na}$, $^{30}_{15}\text{P}$, $^{34}_{17}\text{Cl}$ and $^{38}_{19}\text{K}$ —though apart from the Wigner rules the case of $^{30}_{15}\text{P}$ might be regarded as that of an ordinarily allowed transition (see figure 3).

Because practically all species of the type $\begin{pmatrix} 2Z \\ Z-1 \end{pmatrix}$ with $Z < 22$ are stable species, examples of super-allowed transitions are much less frequent amongst negative-electron-active bodies than with positron-emitters. In fact ^3_1H is the only representative of the first Wigner class, and ^6_2He , $^{10}_4\text{Be}$ and $^{14}_6\text{C}$ are the sole representatives of the second class, amongst the radioelements for which figure 2 has been drawn. As with the positron-emitters, ^3_1H , as a class I body, undergoes a transition which empirically appears super-allowed †, whilst ^6_2He of class II is, but $^{10}_4\text{Be}$ and $^{14}_6\text{C}$ are not, in the super-allowed group according to the diagram. In this connection it may be mentioned that although the very varied behaviour of the different class II species has yet to receive a satisfactory explanation on any theory, the fact that super-allowed transitions occur with some species of this class is unambiguous proof that for these species the Fermi-type interactions are inadmissible (Konopinski 1943, Wigner and Feenberg 1942). A final obvious remark is that the position of the point for $^{12}_5\text{B}$ (figure 2) is entirely anomalous according to the Wigner rules.

Summing up, we may say that the Wigner rules claim evident success in describing the otherwise inexplicably short lifetimes of the $\begin{pmatrix} 2Z \\ Z-1 \end{pmatrix}$ positron emitters, but to be set against this is the very considerable concentration of points about what appears to be the ordinary allowed line A for negative-electron-active species in figure 2. There would appear to be some consideration which restricts the effect of these rules even in the class of the lightest elements to which they should most clearly apply. In this connection the exact lie of the line A of figure 2 is of some importance. This line has been drawn purely empirically as the best straight line having regard to the experimental points. As transferred to figure 3 (A') it is seen to represent equally closely the slope of the super-allowed line S. Furthermore it may be shown to be the straight line which best approximates to the theoretical predictions of the Fermi theory. The dotted line F in figure 2 gives the Fermi function $f(0, W_0)$ (Konopinski 1943) fitted to A in the region of $E = 1$ mev. The theoretical line obviously appears as in many ways an even better representation of the experimental results than is our empirical line A;

* The choice of this particular range of Z for figures 2 and 3 was made precisely because no species belonging to the first Wigner class are at present known for $Z > 21$.

† The β -disintegration of the neutron should also be super-allowed: in this case interpolation from figure 2 would indicate a half-value period of 7 ± 1 minutes.

thus the well-determined point for ^{35}S is no longer significantly out of place, and the super-allowed points for ^3H and ^6He are more nearly the same distance above the line. Also, the anomaly regarding the point for ^{12}B is clearly reduced.

Having established the significance of the allowed line A in this way by comparison with theory, it is of interest to discuss the exact lie of the empirically-drawn allowed line O of figure 1 in a similar manner. We can do this most easily by reference to line F, figure 2, which provides the normalization for the theoretical results. In the table we give the difference of the ordinates ($\Delta = \log \lambda(\text{O}) - \log \lambda(\text{F})$, for E constant) of corresponding points on the two lines as a function of the disintegration energy E . The theoretical differences, $\Delta'(81)$, $\Delta'(93)$, defined by $\Delta'(Z) = \log[f(Z, W_0)] - \log[f(0, W_0)]$ (see Konopinski 1943) are given for comparison*. In the last line of the table the quantity δ is the second difference

Table

$E(\text{mev.})$	0.0158	0.0511	0.158	0.307	0.511	1.022	1.533	2.044
$\log E$	$\bar{2}.20$	$\bar{2}.708$	1.20	$\bar{1}.487$	$\bar{1}.708$	0.009	0.185	0.310
Δ	1.47	1.47	1.32	1.16	1.13	0.97	0.84	0.69
$\Delta'(81)$	2.52	2.23	1.82	1.51	1.45	1.27	1.20	1.10
$\Delta'(93)$	2.91	2.58	2.14	1.81	1.72	1.51	1.42	1.30
δ	1.25	0.94	0.66	0.50	0.46	0.42	0.47	0.51

$\frac{1}{2}[\Delta'(81) + \Delta'(93)] - \Delta$. The constancy of δ over the range $0.2 \text{ mev.} < E < 2.0 \text{ mev.}$ indicates that the straight line O, figure 1, has the full support of theory as representing the variation of λ with E over the range in question. This conclusion may be stated conversely and expanded as follows: the Fermi theory of allowed transitions appears to be entirely adequate so far as predictions of the dependence of disintegration constant on energy and charge number are concerned. The latter point is of interest, and the numerical value of δ over the range of E for which it is constant indicates that allowed transitions of the heaviest elements are less probable than the transitions which we have now with some confidence identified as "ordinarily allowed" for the lightest elements (see lines A and F, figure 2) only by a factor of roughly 3 (for the same energy release). This is of importance in relation to the problem of the extent to which the wave-functions of initial and final nuclear states "overlap" for nuclei of such different degrees of complexity.

The increase of δ for $E < 0.2 \text{ mev.}$ which is evident from the table implies that the theoretical allowed line for the heaviest elements lies above the empirical straight line O of figure 1 at the low-energy end, the deviation being somewhat similar to the deviation of the theoretical line F from the straight line A of figure 2. This conclusion is not surprising and does not invalidate any of our previous discussions, for the purpose of which the empirical allowed line has been taken as reference line. Only with the heaviest elements it should be pointed out that for the extreme low-energy end of the line ($E < 0.02 \text{ mev.}$) the predictions of the simple theory are obviously inadequate, since they do not take count of the energy derived from the "adiabatic" reorganization of the extranuclear electrons which for these elements is of the order of 0.01 mev.

* In this calculation we have falsified theory only to the extent of using p_0 , the maximum momentum, instead of p , an undefined mean momentum which enters as a slowly-varying term in the expression for $f(Z, W_0)$.

There is just one feature of figures 2 and 3 which remains to be considered. We have so far concentrated attention on the identification of a number of transitions which appear to be allowed according to the spin and parity rules and for which there is no empirical evidence of the intervention of the Wigner rules. It will be obvious, however, that there are many points, particularly on figure 2, which do not come within the scope of this identification. Some of these may refer to transitions on which the Wigner rules impose increased forbiddenness—some of them, therefore, may be allowed points so far as spin and parity changes are concerned—others may represent transitions which are “ordinarily” forbidden to first or higher degree. In this connection, since the condition expressed by (1) is no longer fulfilled, we should not expect to find an ordinary first-forbidden line parallel to the allowed line A. Also, since fairly large relative changes in ρ are involved, as between the lightest and the heaviest radioelements in the range $Z < 22$, there is further reason why the higher-order Sargent lines should not stand out so obviously in the diagram. The question of order of forbiddenness must therefore be settled on other evidence for a number of transitions before much progress can be made in empirical classification.

Partial evidence of one kind is already provided by the experimental results considered in this paper. Figure 2 includes the representative points for 13 radioelements of which the β -disintegration is known to be complex, and with these species some transitions occur which would appear to be ordinarily allowed. In two cases of two-component spectra ($^{25}_{11}\text{Na}$ and $^{34}_{15}\text{P}$) each of the observed components appears to result from such an allowed transition; in six other cases ($^{16}_7\text{N}$, $^{20}_9\text{F}$, $^{27}_{12}\text{Mg}$, $^{28}_{13}\text{Al}$, $^{38}_{17}\text{Cl}$ and $^{41}_{18}\text{A}$) one or more allowed components are present together with another component or components which are to some degree forbidden. It seems reasonable to assume that in these cases the other β -transitions are ordinarily forbidden, that is, that the Wigner rules are of no more effect for them than they are for the allowed transitions with which they compete. If we make this assumption, we can at once identify the representative points numbered 6, 12, 19, 21, 27, 28 and 30 on figure 2 as ordinarily forbidden. On the further likely assumption that some of these points, at least, represent transitions which on spin and parity rules alone are first-forbidden, we then have a general impression of the lie of the second Sargent curve on our diagram*. With this approximate curve in mind we can recognize a further group of points (e.g. 14, 23, 35, 38, 41) which lie between the ordinarily allowed and first-forbidden lines—and we are led to conclude that these belong to species ($^{24}_{11}\text{Na}$, $^{31}_{14}\text{Si}$, $^{45}_{20}\text{Ca}$, $^{46}_{21}\text{Sc}$, $^{48}_{21}\text{Sc}$) for the β -transitions of which the Wigner forbiddenness is greater, and more varied, than it is for the other transitions which we have been considering. It is somewhat surprising to find the heavier species predominant in this group, but this is clearly the division to which our tentative considerations lead.

REFERENCES

- BOTHE, W., and MAIER-LEIBNITZ, H., 1937, *Z. Phys.*, **104**, 604.
BRADT, H., HEINE, H. G., and SCHERRER, P., 1943, *Helv. Phys. Acta*, **16**, 455.
BRADT, H., and SCHERRER, P., 1945, *Helv. Phys. Acta*, **18**, 405; 1946, *Ibid.*, **19**, 307.
CONSTANTINOV, A. A., and LATYSHEV, G. D., 1941, *J. Phys. U.S.S.R.*, **5**, 249.
ELLIS, C. D., 1934, *Internat. Conf. Phys., London*, **1**, 43.

* It is tempting to see this curve as passing through 5, the point for $^{14}_6\text{C}$, for the β -disintegration of which a recent determination of nuclear spin (Jenkins 1948) gives $\Delta I = 1$.

- FEATHER, N., 1948, *Nature, Lond.*, **161**, 431.
 FOWLER, R. H., 1930, *Proc. Roy. Soc. A*, **129**, 1.
 GRAY, L. H., 1937, *Proc. Roy. Soc. A*, **159**, 263.
 GREULING, E., 1942, *Phys. Rev.*, **61**, 568.
 JENKINS, F. A., 1948, *Phys. Rev.*, **73**, 639.
 KONOPINSKI, E. J., 1943, *Rev. Mod. Phys.*, **15**, 209.
 KONOPINSKI, E. J., and UHLENBECK, G. E., 1941, *Phys. Rev.*, **60**, 308.
 LATYSHEV, G. D., 1947, *Rev. Mod. Phys.*, **19**, 132.
 LEE, D. D., and LIBBY, W. F., 1939, *Phys. Rev.*, **55**, 252.
 MARSHAK, R. E., 1942, *Phys. Rev.*, **61**, 431.
 RICHARDSON, H. O. W., 1948, *Nature, Lond.*, **161**, 516.
 RUTHERFORD, Lord, LEWIS, W. B., and BOWDEN, B. V., 1933, *Proc. Roy. Soc. A*, **142**, 347.
 WHEELER, J. A., 1941, *Phys. Rev.*, **59**, 27.
 WIGNER, E. P., 1939, *Phys. Rev.*, **56**, 519.
 WIGNER, E. P., and FEENBERG, E., 1942, *Rep. Prog. Phys.*, **8**, 274.

Investigations using a Permanent-magnet Double β -Ray Spectrograph with Coincidence Counting—I. Experiments with Thorium Active Deposit

By N. FEATHER, J. KYLES AND R. W. PRINGLE

Department of Natural Philosophy, University of Edinburgh

MS. received 27 March 1948

ABSTRACT. A detailed description is given of a semicircular focusing β -particle spectrograph in which a thin source is centrally placed and sheaves of oppositely moving electrons are investigated using two thin-windowed Geiger counters. These counters can be used in a coincidence arrangement, and each is capable of movement so as to accept electrons which have been focused with any radius of curvature between 2.5 cm. and 6 cm.

The arrangement was tested using sources of thorium active deposit. During this work the (coincidence) β -particle spectrum for the ThB.C transition leading to the 238 kev. state of ThC was obtained separately for the first time and its end-point determined as 355 kev. A feeble higher-energy β -component due to ThB was also detected and the end-point estimated at 589 kev.

Redeterminations of the end-points of the β -particle spectra of ThC and ThC'' (by β - γ coincidences) gave values of 2.256 mev. and 1.805 mev., respectively. The intensity of the F-line of ThB was found to be 0.288 photoelectrons per disintegration.

§ 1. INTRODUCTION

THE coincidence counting method of investigating time-correlations in the emission of particles and quanta by unstable nuclei has already become one of the standard techniques in the study of radioactivity (Dunworth 1940, Feather 1940a). A development of this technique, first proposed by Feather (1940b), has now been realized in practice: in the present paper the method will be described in detail and the results which have been obtained in its initial application to the active deposit of thorium will be presented and discussed.

The essential feature of the new method is the study of coincidences between the various components of the particle radiations emitted by a radioactive source, e.g. between the internal conversion electrons (belonging to the β -ray "lines") and the disintegration electrons (of the continuous spectra). For this purpose a special form of semicircular focusing spectrometer is employed, designed so that

(for example) a β -line may be focused on one Geiger counter while another counter is used to scan the continuous spectrum. The spectrum of coincidences will then be the continuous spectrum of those disintegration electrons time-correlated with the photoelectrons of the β -ray line, and thus, by inference, with the γ -ray quanta of corresponding energy.

It is well known that magnetic analysis by semicircular focusing is inconsistent with the use of large solid angles for collection—in our particular arrangement (as shown in § 2) these are of the order of 10^{-3} – 10^{-2} depending upon position; on the other hand, the sensitivity of any coincidence experiment increases rapidly with the collection angles which can be employed. A natural development will therefore be to use the present method with other types of β -particle spectrograph of greater collecting power (see e.g. Roberts, Elliott and Deutsch 1944). Nothing new in principle would necessarily be involved in such a development.

§ 2. THE SPECTROGRAPH

A detailed description of the spectrograph has already been given (Feather 1940 b). The essential features of construction are illustrated in figure 1. Slits S_1 and S_2 select two sheaves of oppositely moving β -particles coming from the source A situated midway between the slits. The two sheaves of electrons are focused by a magnetic field, perpendicular to the plane of the diagram, in the horizontal planes containing respectively the windows of the counters C_1 and C_2 and the defining slits.

The counters can be moved independently in the planes of focusing thus allowing the radius of curvature of either sheaf of electrons to be varied between 2.5 cm. and 6 cm. Electrical connections to each counter are made by two spring strips attached to the counter carriage. One of the strips bears against the side wall of the box and serves to earth the cathode. The other strip makes a sliding contact with a copper wire insulated from the box and supplies the anode voltage (see also figure 8).

A shutter B can be moved down to cover the slit S_1 and so to cut off the electrons from C_1 . Movement of the counters and the shutter is obtained by means of screws which can be rotated from outside the vacuum box by keys turning in ground joints. The inside of the box is covered with aluminium and the slits are bevelled to reduce scattering as much as possible. The box is evacuated to a pressure of 10^{-4} mm. Hg by a three-stage oil diffusion pump and is locked in position in the 5.5 cm. gap between the poles of a large permanent magnet (Cockcroft, Ellis and Kershaw 1932). The use of a permanent magnet allows the magnetic field to be kept constant over the long periods necessary for coincidence counting and the provision of field coils enables the field strength to be set at any value up to 2000 gauss in a few minutes. To avoid opening the box in order to measure the magnetic field a search coil is mounted on the top of the spectrograph box and the change of flux produced when the coil is rotated through 180° in the stray field of the magnet is measured with a fluxmeter. For standardization this change of flux can be compared with that measured with a search coil of known area-turns inside the box. An accuracy of better than 1% can be obtained in this way. A still more accurate value for the effective field strength can be obtained by finding the position of one or other of the prominent β -ray lines in the spectrum of the thorium active deposit. The $H\rho$ values of these lines have been measured with great care (Ellis 1932, Arnoult 1939, Siegbahn 1944) and since the position of

the counters can be determined to 0.01 cm., the effective field strength can be obtained to an accuracy of 0.25% by dividing the accepted value of $H\rho$ by the observed value of ρ . Since several lines can usually be obtained for one field setting, a check on the accuracy of the measurements is generally possible.

The profile of a β -ray line as observed with a constant field setting (figure 2) extends over an appreciable range of ρ . It is necessary to determine which point on the counting-rate-position curve is to be taken as fixing the position of the line, and what correction is necessary in order to obtain the true value of ρ from the value corresponding to the position so fixed. If in figure 3 (which is drawn in perspective) we consider the source Q as emitting electrons of one momentum only, P , then the trajectories in a plane perpendicular to the magnetic field will be circles of radius $\rho_0 = P/He$. An electron from the centre of the source passing through O, the centre of the slit, will arrive at R on the centre line of the plane of focusing at a distance a_0 from O, where $4\rho_0^2 = a_0^2 + b^2$. Taking axes X, Y at R as shown, we may calculate the intensity at a point (x, y) by a method similar to that employed by Li (1937), for a source of dimensions given below.

The variation of intensity with y for fixed values of x is shown for a typical case in figure 4. If we integrate one of these curves with respect to y between the limits $+l$ and $-l$, we obtain the intensity over a narrow strip of width dx and length $2l$ for a fixed value of x , $2l$ being the length of the counter window. This is in effect the solid angle, expressed as a fraction of 4π , subtended by the narrow strip $2ldx$ at the source, for electrons of momentum P . The variation of this solid angle with x is shown in figure 5. If the width of the counter window is w , we can then obtain the solid angle subtended by the entire window at the source (the transmission factor $T(x)$), when the centre of the window is at x , by integrating the curve in figure 5 between the limits $x + w/2$ and $x - w/2$. Figure 6 shows the variation of $T(x)$ with x , or, in other words, the variation of the counting rate for electrons of momentum P as the counter slit is moved in the plane of focusing in a constant magnetic field, i.e. the line-shape of a β -ray line of momentum $P = He\rho_0$. Having regard to figure 6 we have chosen as the point fixing the position a of a line, the point A at half the maximum height on the steep right-hand edge of the counting-rate-position curve, since this point can be located with greater accuracy than the position of the peak. The calculated line-shape then enables us to determine the small correction term (less than 0.25%) which must be subtracted from a in order to obtain a_0 from which the value of ρ_0 is calculated.

The resolving power of the spectrograph, defined as $H\rho/\delta(H\rho)$, where $\delta(H\rho)$ is the change in the $H\rho$ value corresponding to the width of a β -ray line at half the maximum height, is obviously $\rho/\delta\rho$ in the case of a constant-field spectrograph. This ratio can be determined from the calculated line-shape. It is found to change with the radius of curvature, from about 33 for the smallest value of ρ to about 60 for the largest ρ employed. For the setting of the F-line in figure 2 the calculated value of 53.8 may be compared with the experimentally determined value of 56.5, in good agreement with the former.

To map out a continuous spectrum the counting rate in counter (1) is determined for a series of positions of the counter, first with the shutter B open, then with it closed. Differences give the counting rates due to β -rays alone. These differences are corrected for counting-rate losses (usually negligible for the continuous spectrum), and for decay of the source. In order to give a true picture of the continuous (momentum) spectrum, the corrected β -particle rate must

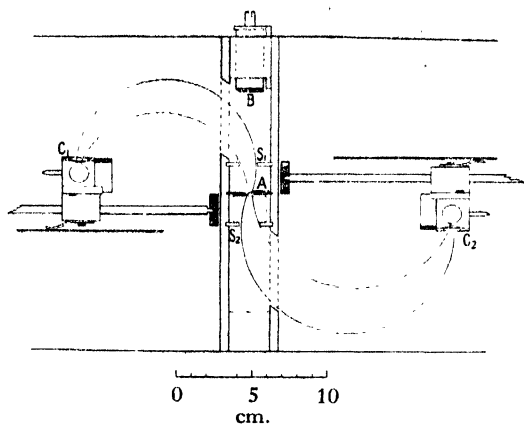


Figure 1. The spectrograph.

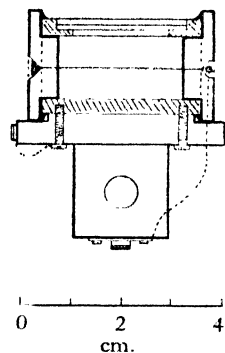


Figure 8. Section through counter.

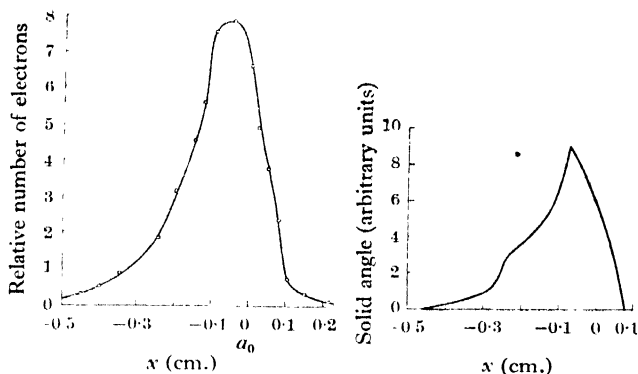


Figure 2.

Figure 5.

Observed shape of ThB F line.
 $H = 237$ gauss. $a_0 = 11.4$ cm.

$b = 2$ cm. $s = 0.4$ cm.
 $2l = 1.9$ cm. $a_0 = 11.4$ cm.

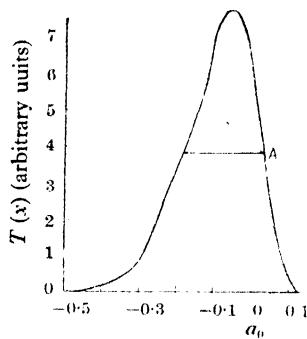


Figure 6.

Variation of $T(x)$ with (x) .
 $a_0 = 11.4$ cm. $b = 2$ cm.
 $\rho_0 = 5.80$ cm. $s = 0.4$ cm.
 $\tau c = 0.10$ cm. $2l = 1.9$ cm.

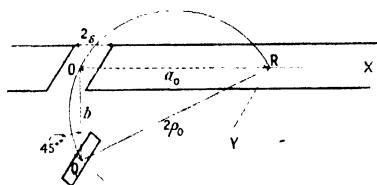


Figure 3.

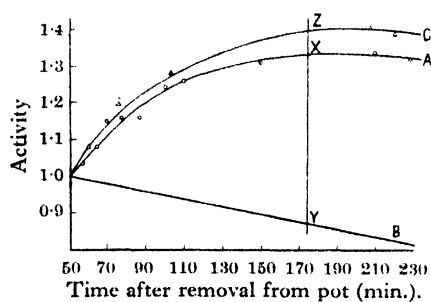


Figure 7.

○ Observed activity at $H\rho = 2500$ gauss-cm.
 Δ " " " $H\rho = 3500$ gauss-cm.

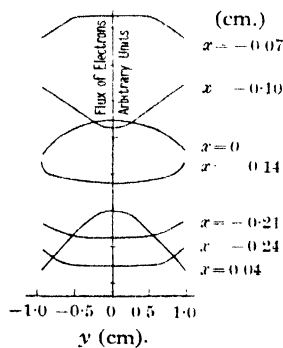


Figure 4.

Source 19×2 mm²
 inclined at 45° .

further be adjusted to refer to the same solid angle of collection and the same interval of $H\rho$. Although this problem might have been treated analytically it was in fact solved empirically.* A value of $H\rho$ in the continuous spectrum was chosen and observations were made with various field strengths $H_1, H_2, \dots$ at radii of curvature $\rho_1, \rho_2, \dots$ such that $H_r\rho_r = \text{constant}$. Let the corrected β -particle rates under these conditions be $N_1, N_2, \dots$. Now, for a constant radius of curvature and a variable field, it is known that the normalizing factor is $1/H$ (Gurney 1925), thus under the conditions of our standardizing experiment the counting rates at $\rho_1, \rho_2, \dots$ for a constant field and a continuous (momentum) spectrum of uniform height would be as $N_1/H_1 : N_2/H_2, \dots$. This means that $H_1/N_1, H_2/N_2, \dots$ are respectively proportional to the normalizing factors corresponding to $\rho_1, \rho_2, \dots$. Plotting these quantities as a function of ρ , a normalizing curve is obtained for the particular counter used. This curve could be checked by the method described in § 6.

Normalization of the continuous spectrum having been effected, it is still necessary for many purposes to know the solid angle for collection to which the spectrum has been normalized. This may be calculated from curves similar to that shown in figure 5. For a radius of curvature of 5.8 cm., for example, and a counter window of area $1.9 \times 0.10 \text{ cm}^2$, we obtain 3.4×10^{-3} of 4π . To be precise, this fraction represents, in the conditions specified, the probability that an electron of the continuous spectrum will enter the counter, the probability being averaged over the momentum range $H\rho$ to $H\rho + \Delta(H\rho)$, where $\Delta\rho$ is the difference between the radii of curvature corresponding to entry by the extreme edges of the counter window. During the experiments described below entrance windows of breadth from 0.10 cm. to 0.15 cm. were used at various times.

The last problem in dealing with the continuous spectrum is that of determining the "representative" ρ value for any setting of the counter. This was solved by a method similar to that used by Lawson and Tyler (1940) for the same purpose.

§ 3. THE SOURCE

Sources of thorium active deposit were obtained by the exposure of aluminium foils of 0.28 mg/cm^2 thickness to the thoron given off by a "dry" emanating preparation of radiothorium, of strength approximately 40 millicuries. For activation a foil is bent over a suitably formed brass block in an arrangement which allows an area of $19 \times 2 \text{ mm}^2$ of foil only to be exposed. Both short- and long-exposure sources were prepared in this way. With long-exposure sources, transient equilibrium having already been attained, the ratio of the absolute activities of ThC and ThB remains constant at $\lambda_C/(\lambda_C - \lambda_B)$ during the subsequent decay; with short-exposure sources this ratio varies with time. The numerical value of the constant ratio is 1.11. Important for our experiment is the fact that with a source in transient equilibrium 1.11 is very closely the ratio of the numbers of electrons in the continuous spectra of Th(C + C') and ThB, respectively, the disintegration constant of ThC' being so much greater than the disintegration constants (λ_B, λ_C) of ThB and ThC.

* Attempts to calculate the normalizing factor in our arrangement are necessarily difficult on account of the form of the counter window (§ 4). In general the electrons do not strike the plane of focusing normally, but arrive at angles to the normal spread over a wider range, and more inclined, the smaller the value of ρ . Thus a fraction of the electrons entering the counter slit (which has finite depth) strikes the far wall of the slit and this fraction increases with decreasing ρ . Some of these electrons, after scattering, may produce a discharge in the counter but many of them will not. Obviously the effect is not susceptible of exact evaluation except by experiment.

With short-exposure sources information concerning the relative contributions of the β -particles of ThB and Th(C + C'') to any given interval of the experimental continuous spectrum can be obtained by the observation of the growth (or decay) of the β -particle counting rate when a source is used which has just been removed from the exposure pot and the Geiger counter is adjusted to accept the appropriate momentum range in the spectrograph. If we consider a source which has been exposed for one hour only, then, whilst the activity due to ThB should decay as shown by curve B, figure 7, the activity due to Th(C + C'') should follow curve C (calculated using standard equations). If an observed activity curve, such as A, lies between the curves C and B then the fraction of the activity due to ThB is given by the ratio XZ/YZ. Clearly, for a particular setting of the counter, this ratio should be constant within the limits of experimental error for any vertical secant to the curves. In order to achieve this result in practice it has been found necessary to wait for about half an hour after the end of an exposure before starting activity measurements of this type. It would appear that during activation atoms of ThC'', given off by recoil from the active deposit which invariably accumulates on the walls of the pot, are collected together with ThB on the foil. If this additional ThC'' activity is not allowed to die out before measurements are begun, the ratio XZ/YZ will be found to vary along the curves.

In general, in order to prevent the escape of recoil atoms from a source in the spectrograph, the activated foil is covered with a similar aluminium foil (0.28 mg/cm² thickness) before being clamped in the source holder. In position, the active deposit, sandwiched between the two thicknesses of foil, is inclined at an angle of 45° to the plane containing the central axes of the two defining slits and is symmetrically placed with respect to this plane. In this connection, the shape of the F-line of ThB was determined first with a bare source and then with a "sandwich" source, but no detectable difference was observed either in the height or shape of the line. It therefore appears that the effect of absorption and straggling produced by the 0.28 mg/cm² covering of the source is negligible—at least for β -particles of energy as great as 147 kev.

§ 4. COUNTERS

The Geiger counters are necessarily of small dimensions since they have to be capable of movement in the restricted space between the magnet pole pieces. While this means that the advantages of a relatively low background counting rate (5–10 min⁻¹) and a short recovery time (30 μ sec.) are enjoyed (Muehlhause and Friedman 1946), at the same time it implies that at high counting rates the life of a counter is short (counting rates as high as 60 000 min⁻¹ had to be sustained for several hours when a counter was registering a strong photoelectron line, or was placed close to the source to count α -particles).

In figures 1 and 8 are shown the details of counter construction. The cathode is made of brass suitably finished to provide a slit 1.0–1.5 mm. broad for the entrance of the β -particles. End pieces are of perspex and are threaded with a central wire of platinum of 0.1 mm. diameter as anode. The thin mica film (~ 1 mg/cm²) which is used to cover the slit is mounted on a concave surface in order to withstand better the atmospheric pressure to which it is subject during the filling process. A small amount of lead screening, 1 cm. thick, is mounted together with the counter on a carriage of ebonite.

A mixture of 10% ethyl alcohol and 90% argon at a pressure of 10 cm. Hg was used to fill the counter after the cathode surface had been washed in alcohol and

polished with silk. The filling line was simply constructed, use being made of rubber connections and vacuum stop-cocks, but the counters were evacuated to 10^{-4} mm. Hg with a diffusion pump. Otherwise the technique employed was similar to that described recently by Regener (1947). Satisfactory "plateaux", 200–300 v. long, were obtained with threshold potentials ranging from 900–1100 v. and positive plateau slopes of about 0.10–0.15% per volt, though this latter quantity was found to increase somewhat when the counting rate was high. Another troublesome effect observed during periods of rapid counting was a gradual increase of threshold voltage (up to 20 or 30 v.) which occasionally made it necessary to use a radium standard to check the counter sensitivity throughout the course of an experiment.

The importance of obtaining an accurate correction for absorption and scattering in the mica window cannot be over-estimated since it becomes very large for energies below 70 kev. ($H\rho \sim 900$ gauss-cm.). Additional thin sheets of mica covering the counter slit were therefore used to plot absorption curves for electrons of a number of energies in the thorium spectrum and in this way, by extrapolation of the results to zero thickness, the correction curve given in figure 9 was obtained. This curve shows reasonable agreement with that of Martin and Townsend (1939), though it should be emphasized that the magnitude of the correction factor for electrons of any energy depends on the precise arrangement of the counter slit and of the mica covering the slit.

§ 5. ELECTRICAL CIRCUITS

In any coincidence experiment it is necessary that provision should be made both for determining the individual rates of entry of particles into the two counters and for counting the coincident pulses which may occur as a consequence of the association in time of these particles.

Figure 10 illustrates the basic circuit used for single channel counting. All H.T. voltage supplies are stabilized. It will be noted that a load resistance of 1 M Ω is used in the counter circuit, but even so, with the small output pulse obtained with a counter of small volume, it is necessary to use a conventional cathode-follower unit mounted close to the counter. The cathode-follower operates through a screened cable into an amplifier designed for the equalization of the pulses (the higher the counting rate the larger the proportion of pulses which are smaller than normal). The particular equalizing circuit used is a modification of the cathode-coupled multivibrator (West 1944). This has proved to be exceedingly stable and reliable in operation.

The 50 k Ω rheostat in the screen grid lead of valve V_2 is used to obtain suitable conditions for operation from triggering pulses, and can in fact be varied over a large range. The output pulses, of 5 μ sec. duration, from the screen grid of V_3 are used to operate a scaling circuit having a factor either of 100 or, for individual rates greater than 50 000 min $^{-1}$, of 3200. The Dynatron Model X200 "Scaler", with its decimal counting system, has proved in every way satisfactory and extremely convenient in use. The input resolving-time of this scaler is 5 μ sec., thus it is evident that the 30 μ sec. recovery time of the Geiger counter alone determines the resolving-time of the counting system as a whole.

It was found that, even at the highest counting rates used (100 000 min $^{-1}$), both the size and shape of the output pulses from the amplifier remained constant. Thus it can be taken as established that the resolving-time for coincidence counting was independent of the single counting rates throughout the entire investigation.

The output to the coincidence mixing set is taken from the anode of V_3 in each amplifier. A conventional Rossi coincidence circuit is used, with a fixed grid bias on the output valve but with a choice of grid leaks ($2000\ \Omega$ and $20\ 000\ \Omega$) for the Rossi valve. In this way two values of the resolving-time, 2.5×10^{-7} sec. and 1.8×10^{-6} sec., are available. The requirement of the highest possible resolving power (smallest τ) has already been stressed (Dunworth 1940). However, attempts to reduce τ much below 2×10^{-7} sec. do not succeed, because of the statistical time-lag between the entry of the ionizing particle and the appearance of the voltage pulse across the counter (Dunworth 1940, Bradt and Scherrer 1943 a); instead genuine coincidences are lost. The larger of the two resolving-times available is used for purposes of checking that there is no loss at the smaller τ and also when delayed coincidences are in question—as when the β -particle of ThC is followed by the α -particle of ThC' after a mean time interval of $(3.7 \pm 0.6) \times 10^{-7}$ sec.—the average life of ThC' (Bradt and Scherrer 1943 b).

A separate recorder was not provided for determining the coincidence rate, although this was only of the order of a few counts per minute; instead, registration on the same scaler as was used for the single counting rates was arranged without inconvenience.

§ 6. EXPERIMENTS WITH THE ACTIVE DEPOSIT OF THORIUM

Thorium active deposit sources were chosen for the initial tests of the new method, and of our spectrograph generally, for two reasons: first because there is present in the complete β -particle spectrum an intense photoelectron line (F-line) eminently suitable for coincidence studies, and secondly because so much information is already available* regarding the radiations in general. In the course of our tests we have, however, been able to add to this information in certain particulars, details of which will be apparent from the descriptions which follow.

(i) Investigations using a single counter

A study was first made of the complete continuous β -ray spectrum of the thorium deposit in transient equilibrium. Because this spectrum extends as far as $H\rho = 9000$ gauss-cm., it was necessary to make measurements at four different values of the magnetic field (250, 500, 1000 and 2000 gauss). From these the distribution curve shown in figure 11 was obtained after the appropriate normalizations (see § 2), corrections for decay and absorption corrections (see § 4) had been made. The fact that a smooth curve resulted from such a series of overlapping junctions provides a good test of the accuracy of the method of normalization.

A further test is provided by a comparison of the areas of the ThB and Th(C + C'') spectra deduced from the curve (the justification for using the dotted projection of the (C + C'') curve to separate the two spectra is given below). These areas, which should be proportional to the numbers of ThB and Th(C + C'') atoms disintegrating per second, were found to be in the ratio 1 : 1.12 in good agreement with the ratio 1 : 1.11 quoted in § 3.

The continuous β -spectrum of the ThB having been obtained, the absolute intensity of the strong F-line of ThB may be calculated. The area under the F-line profile (§ 2, figure 2) represents the number of photoelectrons emitted per

* Gurney 1926, Ellis 1932, Sargent 1932, Sze Shin Yuan 1933, Henderson 1934, Wang 1934, Surugue 1937, Arnoult 1939, Flammersfeld 1939, Siegbahn 1944, Siegbahn and Johansson 1946.

second through a known solid angle, hence, adjusting the solid angle for collection to be equal to that assumed in normalizing the continuous spectrum, we have number of F-electrons per disintegration = $\frac{\text{area under F-line profile}}{\text{area of } \beta\text{-spectrum}} = 0.288$.

This result agrees more closely with that of Flammersfeld (0.284)* than with the results of Arnould (0.222) or Gurney (0.250).

For the end-point of the ThC spectrum a Fermi plot of the single counting rates for $H\rho > 7500$ is given in figure 12, where $(N/F)^{\frac{1}{2}}$ is plotted with F calculated for $Z = 83$. From the point of intersection with the energy axis we obtain $E_{\max} = 2.256$ mev. ($H\rho = 9070$ gauss-cm.), in good agreement with the result of Henderson (1934), $H\rho = 9040$.

The end-point of the ThB spectrum cannot similarly be determined by simple extrapolation. It will be shown in (ii) below how the main component of this spectrum (associated with the γ_F transition) can be separated out. However, it appeared from figure 11 that a component of higher energy might also be present, and experiments with a single counter were made to test this assumption. A series of measurements was performed of the growth and decay of the β -activity at different points of the spectrum, using sources exposed for one hour only. We have already discussed (§ 3) the way in which the activity of such a "non-equilibrium" source should vary after a "short" exposure (see figure 7).

With the counter at a setting corresponding to $H\rho = 3500$ gauss-cm., experimental points were obtained showing the presence of a pure Th(C + C'') activity (figure 7, C). At $H\rho = 2500$ gauss-cm., on the other hand, the growth of activity was less marked (figure 7, A), indicating the presence after 50 minutes of an effect due to ThB amounting to $12 \pm 3\%$ of the total. It can be shown that this effect would correspond to $7 \pm 2\%$ of ThB activity once transient equilibrium had been reached (after 50 minutes the ThC in the source is still growing, while the ThB is of course decaying in amount). Thus in figure 11 we can identify the point X with certainty as belonging to the (C + C'') spectrum. To complete the low energy end of this spectrum ($H\rho < 3000$ gauss-cm.) the distribution given by Flammersfeld (1939) has been used.

We have just seen that the ThB spectrum must have a component extending beyond $H\rho = 2300$ gauss-cm., the currently accepted end-point. It is not difficult to imagine that this weak high energy component, with apparent end-point about 589 kev., corresponds to a ground state-ground state transition between ThB and ThC. The maximum energy of the partial component associated with the γ_F transition is shown in (ii) to be 355 kev., and the energy of the F γ -ray (emission of which is believed to lead to complete de-excitation of the nucleus) is 238.2 kev. The numerical check ($355 + 238 \sim 589$) is fairly convincing.†

* Flammersfeld's value 0.284 was deduced on the assumption that the areas of the continuous spectra of ThB and Th(C + C'') should be equal. He apparently did not regard the difference of 3.5% in area which he found as significant. If his source was in transient equilibrium there should have been a difference in area of 11% as stated above, and in that case values of 0.289 and 0.310 would have been obtained if he had referred his F-line intensity to the observed areas which he assigned to ThB and Th(C + C'') respectively.

† We have to thank Dr. H. O. W. Richardson of this department for drawing our attention to his work on the ThB spectrum. His results, as yet unpublished, and those of Flammersfeld (1939) both clearly show the existence of the high-energy component, though it is remarkable that Flammersfeld does not comment upon this feature of his curve.

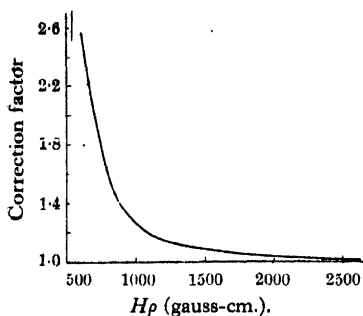


Figure 9. Correction curve for mica window absorption (1.3 mg/cm^2).

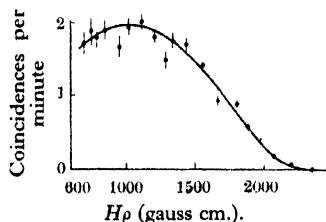


Figure 13. Spectrum of coincidences of disintegration electrons with electrons of the F-line. The vertical lines represent probable errors.

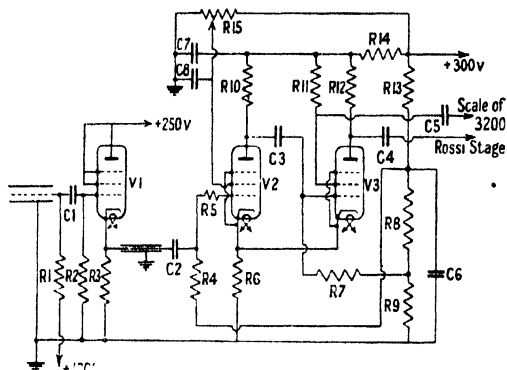


Figure 10. Single channel circuit.

COMPONENTS

Resistances

R1 $1 \times 10^6 \Omega$	R6 $7.5 \times 10^3 \Omega$	R11 $10 \times 10^3 \Omega$
R2 0.25×10^6	R7 100×10^3	R12 2×10^3
R3 10×10^3	R8 320×10^3	R13 220×10^3
R4 50×10^3	R9 100×10^3	R14 5×10^3
R5 100	R10 10×10^3	R15 50×10^3

Condensors

C1 $100 \mu\text{F}$	C5 $1000 \mu\text{F}$
C2 $150 \mu\text{F}$	C6 $8 \mu\text{F}$
C3 $25 \mu\text{F}$	C7 $8 \mu\text{F}$
C4 $50 \mu\text{F}$	C8 $8 \mu\text{F}$

Valves

V1	EF 50
V2	
V3	

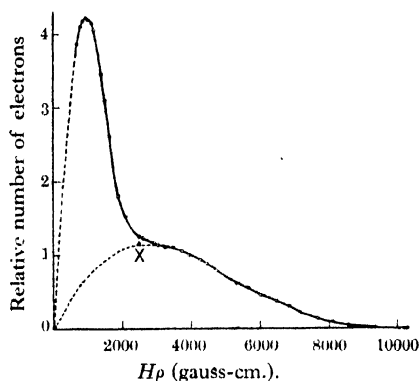


Figure 11. The momentum distribution of the disintegration electrons of the active deposit of thorium in transient equilibrium. Ordinates normalized to equal momentum intervals and corrected for counter window absorption.

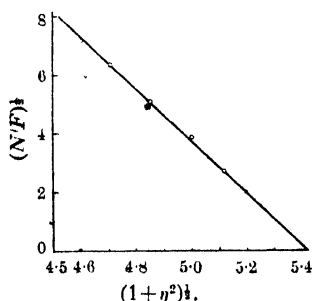


Figure 12. Fermi plot giving ThC end point.

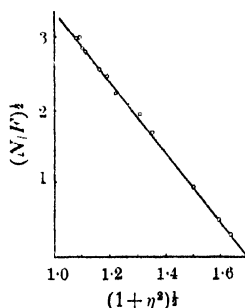


Figure 14. Fermi plot of points given in figure 13. F calculated for $Z=82$.

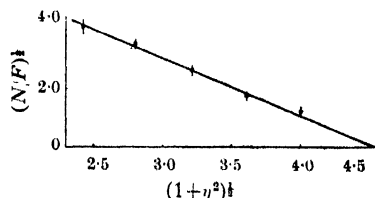


Figure 15. Fermi plot of β - γ coincidences (ThC^{*} spectrum). Probable errors are shown.

(ii) *Investigations using the coincidence technique*

β -photoelectron coincidences. The number of genuine coincidences to be expected between photoelectrons entering counter (1), when it is centred on a β -line, and disintegration electrons entering counter (2), as it is moved through the continuous spectrum, may be calculated as to order of magnitude as follows.

Let photoelectrons enter counter (1) at the rate of $N_p \text{ min}^{-1}$. Then of the corresponding disintegration electrons a fraction $\omega_2 F(P) \Delta P$ will enter counter (2) where $4\pi\omega_2$ is the solid angle subtended by the counter window at the source, $F(P)$ is the momentum distribution function describing the partial spectrum for the disintegration mode with which the β -line is associated, and ΔP is the range in momentum, about the mean momentum P , of the particles which enter the counter. The value of ω_2 , as we have seen in § 2, may be assumed to be of the order of 4×10^{-3} , and since $F(P)$ must be normalized so that $\int_0^{P_m} F(P) dP = 1$, P_m being the upper limit of the (partial) momentum spectrum, $F(P) \Delta P$ may be of the order of $1/50$. The coincidence rate to be expected is then $N_G = N_p \omega_2 F(P) \Delta P \sim 4 \text{ min}^{-1}$, if we take N_p as $50\,000 \text{ min}^{-1}$. This value of N_p could conveniently be attained in our arrangement for the F-line of ThB.

When counter (1) was fixed on this line, coincidences were observed with counter (2) at various positions in the continuous spectrum. A coincidence resolving-time of $2.5 \times 10^{-7} \text{ sec.}$ was used, and a typical set of results obtained near the momentum peak of the main ThB spectrum reads :

N_2 (continuous spectrum + γ background)	$= 7000 \text{ min}^{-1}$
N_1 (F-line + continuous spectrum + γ background)	$= 40000 \text{ min}^{-1}$
N_{12} (coincidences observed)	$= 5.1 \text{ min}^{-1}$
N_A (accidental coincidences $= 2N_1N_2\tau$)	$= 2.3 \text{ min}^{-1}$
N_G (genuine coincidences)	$= 2.8 \text{ min}^{-1}$

It may be noted that with a counting period of one hour the accuracy of determination of N_G is only about 8% in this case.

An extended series of observations was carried out with the object of deducing the complete momentum spectrum of coincidences. The usual corrections were made—for decay, for absorption in the counter window, and for the variation of the effective solid angle of collection with the position of the counter (see § 2)—as also corrections for (γ - γ) and (β - γ) coincidences ($< 5\%$ of the total even in the most unfavourable case). There was no necessity to correct for cosmic-ray coincidences, which were negligible throughout. After all corrections had been made, the final spectrum of coincidences was as shown in figure 13. Two field strengths ($H = 280$ and 440 gauss) were used to obtain this spectrum.

No genuine coincidences with the F-line were detected beyond $H\rho = 2350$ gauss-cm., and the observed points were found to fit a Fermi plot ($Z = 82$) very well (figure 14). The Fermi end-point appears at $(1 + \eta^2)^{\frac{1}{2}} = 1.695$ or $E_{\max} = 355 \text{ kev.}$, the value used above.

An increase of resolving-time to $1.8 \times 10^{-6} \text{ sec.}$ was found to be without effect on the number of genuine coincidences observed, thus we may conclude that the lifetime of the 238 kev. excited state of ThC cannot be greater than 10^{-7} sec. at the most.

While these results are to be regarded as satisfactory, the method obviously allows of much improvement in the direction of shortening the periods of counting which are required under conditions of great stability of operation of counters.

and circuits. A limiting factor at present is undoubtedly in the Geiger counter itself, the essential characteristics of which determine the smallest resolving-time which can profitably be employed. If the electron multiplier could be used in place of the counter, the scope of the method would clearly be greatly extended; this change would probably enable τ to be reduced by a factor of 10^2 and the loss of accuracy involved in the background of accidental coincidences to be largely avoided.

β - γ coincidences. In an experiment similar to that of Bothe and Maier-Leibnitz (1937) one of the counters, shielded so as to record only γ -rays, was placed as near as possible to the source (dotted position in figure 1) and another counter used to traverse the continuous spectrum as before. Coincidences were observed for electron energies above 650 kev. ($H\rho > 3500$ gauss-cm.).

Since the disintegration electrons of ThC are followed by only a small amount of medium-energy γ -radiation, the spectrum of coincidences which is obtained in this arrangement will follow very closely the β -spectrum of ThC'' (which emits a 2.62 mev. γ -ray in every disintegration). The Fermi plot (figure 15) of the experimentally determined coincidence spectrum is a straight line cutting the axis at $E_{\max} = 1.805 \pm 0.02$ mev.* Henderson (1934) obtained the value 1.795 mev. using a pure ThC'' source by a single counting method in a magnetic spectrograph; the only other published determination with such a source is that of Chalmers (1929) who obtained 1.75 mev. by an absorption method.† The coincidence resolving-time used in our experiments was 2.5×10^{-7} sec.

α - β coincidences. The α -particles from ThC' (8.6 cm. range) follow the disintegration electrons from ThC at a mean interval of $(3.7 \pm 0.6) \times 10^{-7}$ sec. (see above). Thus if an α -particle counter is set up near the source, with an 8 mg/cm² sheet of mica (to cut out the unwanted α -particles from ThC of 4.7 cm. range) replacing the lead screen used in the last arrangement, it will register coincidences with the disintegration electrons of ThC if the longer resolving-time (1.8×10^{-6} sec.) is used. In this way the total β -spectrum of ThC may be mapped out once the contribution of the β - β ‡ and γ - β coincidences has been assessed ($< 10\%$ of the total). Because of the increased resolving time the results obtained with the present arrangement were not good enough to reproduce accurately the shape of this spectrum, but they established the fact that genuine α - β coincidences could be recorded up to the known end-point (9070 gauss-cm.).

Photoelectron-photoelectron coincidences. The β -spectrum of the thorium deposit contains no suitable pair of lines for the study of a cascade process, but the work is being continued with the active deposit of radium in the β -spectrum of which there are a number of β -lines (e.g. A, B, F, G and H) all of sufficient intensity to allow of such an investigation. This work will be reported on in a later communication.

ACKNOWLEDGMENTS

The permanent magnet and associated equipment used in this research was originally purchased by means of grants from the Caird Fund and the Government

* Quite recently Siegbahn and Johansson (1946) have reported measurements by the method described in this section and have given a value of 1.72 mev. for the ThC'' end-point.

† The quoted value has been deduced from Chalmers' measured range using the currently accepted range-energy relation.

‡ The use of a slit system to separate α - and β -particles magnetically proved to be a disadvantage because of the deterioration in the solid angle for collection.

Grants Committee of the Royal Society, and was used in the Cavendish Laboratory, first by Professor C. D. Ellis, F.R.S., and his associates, and later in work directed by one of us (N.F.). We have to thank the Royal Society and Sir Charles Ellis, as also the Director of the Cavendish Laboratory, for their agreement that the equipment should be moved to Edinburgh and used in this investigation. One of us (R.W.P.) has been in receipt of a Senior Research Award from the Department of Scientific and Industrial Research during the later stages of this research and wishes to express his appreciation of the opportunities so afforded.

REFERENCES

- ARNOULT, R., 1939, *Ann. Phys., Paris*, **12**, 241.
 BOTHE, W., and MAIER-LEIBNITZ, H., 1937, *Z. Phys.*, **104**, 604.
 BRADT, H., and SCHERRER, P., 1943 a, *Helv. Phys. Acta*, **16**, 251; 1943 b, *Ibid.*, **16**, 259.
 CHALMERS, J. A., 1929, *Proc. Camb. Phil. Soc.*, **25**, 331.
 COCKCROFT, J. D., ELLIS, C. D., and KERSHAW, H., 1932, *Proc. Roy. Soc. A*, **135**, 628.
 DUNWORTH, J. V., 1940, *Rev. Sci. Instrum.*, **11**, 167 (containing a list of the earlier papers).
 ELLIS, C. D., 1932, *Proc. Roy. Soc. A*, **138**, 318.
 FEATHER, N., 1940 a, *Rep. Prog. Phys.*, **7**, 66; 1940 b, *Proc. Camb. Phil. Soc.*, **36**, 224.
 FLAMMERSFELD, A., 1939, *Z. Phys.*, **114**, 227.
 GURNEY, R. W., 1925, *Proc. Roy. Soc. A*, **109**, 540; 1926, *Ibid.*, **112**, 380.
 HENDERSON, W. J., 1934, *Proc. Roy. Soc. A*, **147**, 572.
 LAWSON, J. L., and TYLER, A. W., 1940, *Rev. Sci. Instrum.*, **11**, 6.
 LI, K. T., 1937, *Proc. Camb. Phil. Soc.*, **33**, 164.
 MARTIN, L. H., and TOWNSEND, A. A., 1939, *Proc. Roy. Soc. A*, **170**, 190.
 MUEHLHAUSE, C. O., and FRIEDMAN, H., 1946, *Rev. Sci. Instrum.*, **17**, 5c6.
 REGENER, V. H., 1947, *Rev. Sci. Instrum.*, **18**, 267.
 ROBERTS, A., ELLIOTT, L. G., and DEUTSCH, M., 1944, *Rev. Sci. Instrum.*, **15**, 178.
 SARGENT, B. W., 1932, *Proc. Roy. Soc. A*, **139**, 659.
 SIEGBAHN, K., 1944, *Ark. Math. Astr. Fys.*, **30 A**, 20.
 SIEGBAHN, K., and JOHANSSON, A., 1946, *Ark. Mat. Astr. Fys.*, **34 A**, 10.
 SURUGUE, J., 1937, *Ann. Phys., Paris*, **8**, 484.
 SZE SHIN YUAN, 1933, *Ann. Phys., Paris*, **19**, 59.
 WANG, K. C., 1934, *Z. Phys.*, **87**, 633.
 WEST, S. S., 1944, *Electronic Engng.*, **17**, 118.

Note on the Impedance Variations of an Electro-Acoustic Transducer in a Reflecting Field

By S. BYARD

Royal Naval Scientific Service, Department of Physical Research, Admiralty

Communicated by A. B. Wood; MS. received 2 March 1948

WHEN a reflecting surface is introduced into the field of an electro-acoustic transducer so that a certain proportion of the radiated energy is returned to the source, the measured electrical impedance of the transducer will vary according to the relative phasing of the returned energy. Consequently if the reflecting surface be moved towards or away from the transducer, or alternatively, if the frequency of the electrical input be varied, periodic fluctuations may be observed in the input voltage or current. The Pierce interferometer uses this effect in determining acoustic wavelength in a liquid or gas. The purpose of this note is to relate the impedance changes responsible for the voltage and current fluctuations to the efficiency of the electro-acoustic device by making use of the

reciprocity principle. It is also indicated how, in theory at least, the absolute calibration of a projector or a receiver may be obtained by impedance measurements on a single instrument, and how this is related to Carstensen's more practical "self-reciprocity" method of determining transducer constants.

Let E_0 be the open-circuit voltage response of an electro-acoustic transducer when placed in a sound field of pressure $P e^{-i\omega t}$. If M is the receiver sensitivity expressed in open-circuit volts/dyne/cm², then

$$E_0 = MP. \quad \dots\dots(1)$$

Let Z_0 be the nominal electrical impedance of the transducer. Suppose now that a voltage E , of the same frequency as E_0 , be applied to the transducer terminals; the equivalent circuit is shown in figure 1, and the resulting current I is given by the expression

$$E = IZ_0 + E_0. \quad \dots\dots(2)$$

The measured impedance Z under these conditions would be

$$Z = E/I = Z_0 + E_0/I, \quad \dots\dots(3)$$

or from (1), expressing E_0 in terms of P ,

$$Z = Z_0 + MP/I. \quad \dots\dots(4)$$

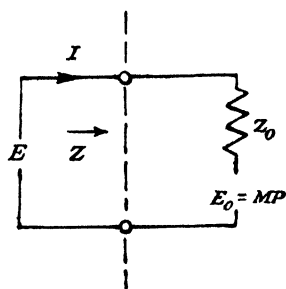


Figure 1.

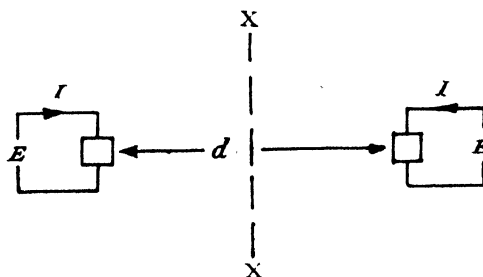


Figure 2.

The field P may be assumed to be produced by an identical transducer excited by a current I at the same frequency, and situated at a distance d from the first transducer as in figure 2. If S_d is the pressure produced at this distance d per ampere excitation, then

$$P = S_d I, \quad \dots\dots(5)$$

and substituting this value for P in (4),

$$Z = Z_0 + MS_d. \quad \dots\dots(6)$$

The factor MS_d in (6) above is the change in impedance due to the sound field of the second transducer and is added vectorially to Z_0 in order to obtain Z . Z will have a maximum whenever MS_d is in phase with Z_0 , and a minimum when MS_d is 180° out of phase, and the impedance will show periodic fluctuations of diminishing amplitude as the transducers are moved away from one another.

Let two impedance measurements Z_1 and Z_2 be made at an adjacent maximum and minimum when the transducers are at distances d_1 and d_2 apart respectively ($d_2 = d_1 + \lambda/2$, where λ is the wavelength). Then if S_1 and S_2 be the appropriate transmitting constants obtaining at the two positions,

$$|Z_1| = |Z_0| + |MS_1|; \quad |Z_2| = |Z_0| - |MS_2|, \quad \dots\dots(7)$$

and the difference between these two moduli is

$$|\Delta Z| = |Z_1| - |Z_2| = |M(S_1 + S_2)|. \quad \dots\dots(8)$$

The reciprocity parameter J which relates the transmitting and receiving responses of a transducer is given (Schottky 1926, Maclean 1940) by

$$J = |M/S_a| = (2\lambda d/R) \times 10^{-7}, \quad \dots\dots(9)$$

where λ is the wavelength of the sound and R the characteristic acoustic impedance of the medium in c.g.s. units. If this is assumed to be valid for the system, it follows that

$$|\Delta Z| = \left| \frac{M^2 R}{2\lambda} \left(\frac{1}{d_1} + \frac{1}{d_2} \right) \right| \times 10^7. \quad \dots\dots(10)$$

Now suppose the second transducer replaced by a perfect reflector, XX in figure 2, which can be moved to satisfy the required impedance conditions. If d' and d'' are the two distances of the reflector from the source, satisfying the two impedance conditions, then equation (10) becomes

$$|\Delta Z| = \left| \frac{M^2 R}{2\lambda} \left(\frac{1}{2d'} + \frac{1}{2d''} \right) \right| \times 10^7 \simeq \left| \frac{M^2 R}{2\lambda d} \right| \times 10^7, \quad \dots\dots(11)$$

where d is the mean of d' and d'' . In this equation the distances between source and reflector are doubled to give the total lengths (d_1 and d_2 in (10)), of the transmission paths. The approximation is permissible when $d \gg \lambda$, and this will obtain at those distances where the reciprocity parameter defined in (9) is valid, and where the inverse relationship between pressure and distance is established.

From (9) and (11), the transducer constants may be expressed in terms of ΔZ , R , λ and d :

$$|M| = [(2\lambda d/R) |\Delta Z| \times 10^{-7}]^{\frac{1}{2}}; \quad |S_a| = [(R/2\lambda d) |\Delta Z| \times 10^7]^{\frac{1}{2}}. \quad \dots\dots(12)$$

An examination of these expressions with reference to the performance of practical transducers shows that their application to the absolute calibration of a transducer would be limited to efficient projectors functioning at resonance, and that in any case the accuracy will be low since the result depends upon the measurement of a small change in impedance. Carstensen (1947) has, however, recently described a pulse technique in which the transmitter current I and the received voltage E_0 are measured separately. A train of waves is used short enough to ensure that the transmission does not overlap the reception of the echo. Reference should be made to the original paper for practical details and for a discussion of the limitations. The expressions for the transducer constants $|M|$ and $|S_a|$ are, in terms of E_0 and I ,

$$|M| = [(4\lambda d/R)(E_0/I) \times 10^{-7}]^{\frac{1}{2}}; \quad |S_a| = [(R/\lambda d)(E_0/I) \times 10^7]^{\frac{1}{2}} \quad \dots\dots(13)$$

and it will be observed that the ratio E_0/I in these expressions is the added impedance of equation (3) and is equal to half the differential impedance ΔZ .

Thanks are due to the Admiralty for permission to publish this note.

REFERENCES

- CARSTENSEN, E. L., 1947, *J. Acoust. Soc. Amer.*, **19**, 961.
 MACLEAN, W. A., 1940, *J. Acoust. Soc. Amer.*, **12**, 140.
 SCHOTTKY, W., 1926, *Z. Phys.*, **36**, 689.

LETTERS TO THE EDITOR

Dependence of β -Decay on the Sign of the Change in Nuclear Spin

The calculation of the probability of β -decay for given values of the magnitudes of the initial and final angular momenta of the nucleus may be abbreviated, as has been done by Marshak (1942), by application of a sum rule and of conservation of angular momentum. Professor Feather and Dr. Richardson have pointed out to me a discrepancy between Marshak's results and a statement of Greuling (1942). Here Marshak's argument is amplified and Greuling's statement shown to be wrong.

Under certain approximations the interaction between the nucleons and the electron-neutrino field contains terms of the type V_N, V_F , where V_N, V_F are respectively functions of the nuclear variables only, and of the field variables only. Let $P(j, m : j', m')$ denote the probability of transition of the system from a state where the angular momentum of the nucleus corresponds to the usual designation j, m to the state labelled by j', m' , the probability being summed for all consistent values of energy, linear momenta, angular momenta and spin momenta of the electron-neutrino field. Total angular momentum is conserved to the usual approximation. Since V_N is Hermitian, and since the part of the transition probability arising from V_F is independent of the sense of the change of angular momentum,

$$P(j, m : j', m') = P(j', m' : j, m). \quad \dots\dots (1)$$

Define $Q(j, j')$ to be the probability of the transition from a single m belonging to j to any of the $2j+1$ values of m' belonging to j' : thus

$$Q(j, j') = \sum_{m'=-j'}^{j'} P(j, m : j', m'). \quad \dots\dots (2)$$

That $Q(j, j')$ is, as the notation suggests, independent of m , may be seen as follows. The usual expansion of the light particle wave functions (evaluated at the position of a nucleon) in ascending powers of the radius (see for example Konopinski and Uhlenbeck 1941) means that eventually the operator V_N contains, among other things, coordinate vectors of the nucleons: if ψ_e, ψ_n are wave functions of electron, neutrino the term $\psi_e^* V_F \psi_n$ contains operators (in V_F) not involving spatial coordinates, together with spherical harmonics, etc. When an integration over the azimuth ϕ is performed, the integrand contains the azimuth in a factor of the form $\exp(iM\phi)$ and in terms involving coordinate vectors of the nucleons. This is now entirely similar to the theory of light emission, and a similar sum rule (see for example Condon and Shortley 1935) is valid.

Summation of (2) for all values of m belonging to j gives, on account of (1), the usual result:

$$(2j+1)Q(j, j') = (2j'+1)Q(j', j). \quad \dots\dots (3)$$

Marshak's simplification. Let k, μ specify the angular momentum of the electron-neutrino field so that, using suffixes i, f to denote initial, final values,

$$|j_i - j_f| = k, \quad \dots\dots (4)$$

$$m_i - m_f = \mu. \quad \dots\dots (5)$$

Suppose first that $j_i - j_f > 0$. We may choose $m_i = j_i$ in evaluating (2). Thus

$$j_i - j_f = k \geq \mu = m_i - m_f = j_i - m_f.$$

Hence $j_f \leq m_f$ and, since also $j_f \geq m_f$, we must have $j_f = m_f$. Thus only one term occurs in the summation (2), viz.

$$Q(j_i, j_f) = P(j_i, j_i : j_f, j_f). \quad \dots\dots (6)$$

Then, for $j_i < j_f$, (3), (6) give $Q(j_i, j_f) = [(2j_f+1)/(2j_i+1)]P(j_f, j_f : j_i, j_i)$.

The discrepancy. Greuling's equations (1), (3) contain a summation over m which corresponds to our summation over μ or m_f . He does not explain in detail his derivation of his equations (11), but the general implication of his statement in the second paragraph of p. 575 of his paper seems to be wrong. For $j_i > j_f$ Greuling's summation over m should be equivalent to Marshak's result. Marshak's summation reduces to one term when $m_i = j_i$, but the result would be the same whatever value of m_i were chosen and however many were the terms in the summation. Thus there is no need to multiply half-lives by $2(j_i - j_f) + 1$ if Greuling's equations (1), (3) are correctly applied.

The number of terms in the summation may be easily calculated. By equation (5), for given m_i, j_i, j_f, μ may take the values $m_i - j_f, m_i - j_f + 1, \dots, m_i + j_f$, and, in order that $|\mu| \leq k$, we have by (4) $-|j_i - j_f| \leq m_i - j_f + \lambda \leq |j_i - j_f|$, where λ is any integer satisfying $0 \leq \lambda \leq 2j_f$.

Thus there are $2|j_i - j_f| + 1$ or $|j_i - j_f| - m_i + j_f + 1$ terms according respectively as $-|j_i - j_f| - m_i + j_f \geq 0$ or < 0 . Thus, if $j_i < j_f$, there are $2(j_f - j_i) + 1$ terms, while if $j_i > j_f$ there are $2(j_i - j_f) + 1$ or $j_i - m_i + 1$ terms according as $2j_f \geq$ or $< j_i + m_i$. Thus, if $j_i > j_f$ and $m_i = j_i$, there is just one term.

Department of Natural Philosophy,
University of Aberdeen.
1st September 1948.

C. STRACHAN.

CONDON, E. U., and SHORTLEY, G. H., 1935, *The Theory of Atomic Spectra* (Cambridge: University Press).

GREULING, E., 1942, *Phys. Rev.*, **61**, 568.

KONOPINSKI, E. J., and UHLENBECK, G. E., 1941, *Phys. Rev.*, **60**, 308.

MARSHAK, R. E., 1942, *Phys. Rev.*, **61**, 431.

The Attenuation of Ultra-High Frequency Electromagnetic Radiation by Rocks

We were interested by Mr. Cooper's paper (1948) on the attenuation of very short radio waves in rocks, and would like to draw attention to some very similar work which we carried out at the National Physical Laboratory in 1942, the results of which were published later (McPetrie and Saxton 1945).

We made measurements on wavelengths of 2 and 5 metres of the field strength produced in two specially excavated tunnels by a horizontal aerial fed from a small power transmitter which was erected over each tunnel in turn. The two tunnels were made as nearly identical in structure as possible, except that one was at a depth of 1.15 metres and the other at 1.85 metres. It was hoped that any disturbing "tunnel effect" could be eliminated by making comparative measurements in the two tunnels when subsidiary tests made above ground showed that the same field existed at the surface of the earth immediately above the tunnels. This seems to us a more reliable procedure than that followed by Mr. Cooper, in which an attempt is made to evaluate this tunnel effect theoretically. Any such attempt must necessarily be uncertain. In Mr. Cooper's experiments, fortunately, the transmission path through the earth is sufficiently long for the absorption in the earth to be very much greater than such losses as may be expected, on any reasonable hypothesis, to arise from the effects of the tunnel.

Our two tunnels were close together on the same site, which was of sandy loam overlying gravel, and the observations thus gave directly the attenuation in a path of 0.7 metres through this medium. We also made measurements of the reflection coefficient of the earth's surface on the same site, and these, together with the absorption measurements, enabled us to determine both the dielectric constant and the conductivity of the ground, assuming uniform properties with depth for the latter. Since this assumption is not strictly true, the values so obtained must be regarded only as reasonable mean values. Our results indicated that the attenuation of both 2-metre and 5-metre waves was about 12 db/m., which is not very different from the value of 10.5 db/m. for Pennant sandstone quoted by Mr. Cooper.

The actual figure for the attenuation obtained, as the author states, depends largely upon the amount of water present. We should like to point out, however, that Mr. Cooper's statement that at a frequency of 200 Mc/s. the dielectric losses in water are still negligible in comparison with the conductivity loss is not strictly accurate. In fact, at this frequency for the usual amount of dissolved salts occurring in fresh water, which leads to an ionic conductivity of about 10^8 E.S.U. ($\sim 10^{-13}$ E.M.U.), the dielectric and conductivity losses are just about equal.

Signals Research and Development Establishment,
Somerford, Christchurch, Hants.,

J. S. MCPETRIE.
J. A. SAXTON.

and

Radio Division, National Physical Laboratory,
Teddington, Middlesex.

21st July, 1948.

COOPER, R. I. B., 1948, *Proc. Phys. Soc.*, **61**, 40.

MCPETRIE, J. S., and SAXTON, J. A., 1945, *J. Instn. Elect. Engrs.*, **92**, Part III, 256.

I was very grateful to Drs. McPetrie and Saxton, with whose work I am, of course, familiar, for their comments on my paper. I must apologize for not having referred to them in it, but I should like to point out that my work was concerned with the electrical properties of large thicknesses of rock *in situ*, and not with the properties of the surface soil-layer. I had hoped that the consolidated Pennant Sandstone which I used would have held less water and therefore show a smaller attenuation than the sandy loam mentioned in their work. I agree that their procedure was more reliable, but the experimental difficulties with which I had to contend prevented any more satisfactory arrangements from being made. In fact I find some consolation in the agreement obtained with their results.

As a matter of interest I also observed a rainfall effect similar to that mentioned in their paper during my work; on the last day of the experiments in Pontypridd tunnel rain fell, but it was not until the following day that any diminution was observed in the echo-strength inside the tunnel. This effect was not investigated sufficiently well for inclusion in the paper, however.

Finally, I am indebted to Drs. McPetrie and Saxton for pointing out the error mentioned in the last paragraph of their letter.

Department of Geodesy and Geophysics,
Downing Place, Cambridge.
3rd August 1948.

R. I. B. COOPER.

The Performance of a New Radio-Frequency Ion Source

The observations of Hertzberg (1927) on the low intensity of the molecular spectrum of an electrodeless ring discharge in hydrogen suggested that this method of excitation should prove very effective in achieving a high proton percentage in an ion source. This mode of excitation has been used in a recently described ion source (Bayley and Ward 1948). The characteristics of such an ion source are of great interest to users of accelerated ion beams. The radio-frequency ion source which has been developed in this laboratory is reliable and gives high ion currents, a high atomic content and ions of uniform energy.

A high concentration of atomic hydrogen is maintained in a pyrex discharge tube by reducing the exposed metal surfaces to a minimum and producing a high degree of ionization by a coil round the tube excited by a 20 Mc/s. 150 watt oscillator. The discharge is self-starting provided the probe voltage is off. A tungsten wire anode is isolated from the main discharge by a constriction. This reduces the flow into the discharge space of molecular hydrogen formed by recombination on the anode. The duralumin probe electrode enclosed in a glass cylinder projects into the discharge space as it has been found (Hertzberg 1927) that the concentration of atomic gas and hence the proton percentage diminishes towards the walls. Because of the high concentration of positive ions the exposed probe surface can be reduced to small dimensions and recombination reduced to a minimum.

The form of the electric field resulting from the presence of the glass cylinder around the probe focuses the ions into the canal. This mode of focusing has been demonstrated before (Thonemann 1946).

Figure 1 shows the mechanical details of the source used for making measurements. Approximately 50 watts of high-frequency energy are dissipated in the tube. A small air-blower is used to keep the wax joint cool.

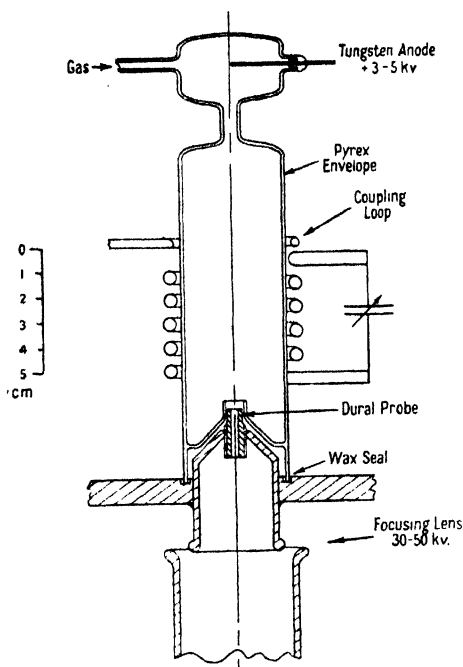


Figure 1. Assembly of the radio-frequency ion source.

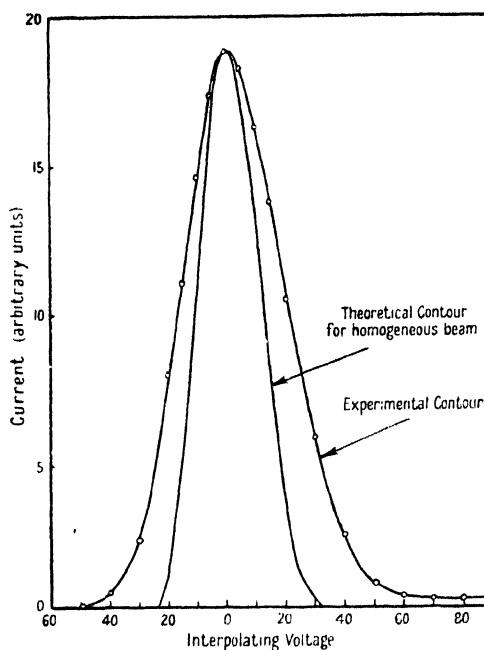


Figure 2. Comparison of the theoretical contour for a homogeneous beam with the D_1^+ ion contour.

$$V_p = 3 \text{ kv.}, V_a = 10 \text{ kv.}, p = 20 \mu; \\ s_1 = 0.1 \text{ mm.}; s_2 = 0.1 \text{ mm.}$$

Table 1 shows the mass spectrum of a deuterium ion beam with about 7% hydrogen present. The various masses and break-up products are easily identifiable. The width of the D_1^+ peak is approximately $\frac{1}{2}$ ma.

Table 1. The mass spectrum of a deuterium ion beam

Magnet current (ma.)	Species	Percentage abundance
29	D_1^+ from dissociation of D_3^+	0.9
32	H_1^+ ; D_1^+ from dissociation of D_2^+	6.0
54	D_1^+ and H_3^+	79.0
60	D_2^+ from dissociation of D_3^+	1.0
62	HD^+ and H_3^+	0.9
80	D_2^+ and H_2D^+	8.8
90	HD_2^+	1.4
99	D_3^+	2.1

The variation of proton and deuteron percentage with pressure in the source is shown in tables 2 and 3. The discharge extinguishes at the low pressure limit. Hydrogen was supplied through a heated palladium tube. The differences in atomic percentage between deuterium and hydrogen may be due to the difference between the radio-frequency powers absorbed in deuterium and in hydrogen.

When a new source is first operated the discharge may change suddenly from its normal brilliant red to a dull white and the proton current may fall by a factor of 10. This is probably due to the evolution of hydrocarbons and other gases from electrodes under positive ion bombardment in the accelerator. After the system has been outgassed the proton current remains stable and shows no deterioration after months of intermittent use.

The homogeneity of the deuterons after a 10 kv. acceleration was measured on a mass spectrometer. A Faraday screen was introduced to eliminate any possible modulation of the ion energy by the electrostatic R.F. field. The discharge had then to be initiated by bringing a Tesla coil close to the upper section of the source. Figure 2 shows the contour for the D_1^+ ion. The theoretical contour from a homogeneous source, taking into account the aberrations in the spectrometer, is given for comparison. It is concluded that 90% of the ions have an energy spread of less than ± 20 electron volts. The magnitude

Table 2

Pressure (microns)	14	16	18	20	23	25	26	28
Percentage	$\left\{ \begin{array}{l} H_1^+ \\ H_2^+ \\ H_3^+ \end{array} \right.$	84	85	88	89	90	90	92
Abundance		14	11	10	8	7	8	5
		2	4	2	3	3	2	3

Table 3

Pressure (microns)	13	15	17	20	22	25	28	30
Percentage	$\left\{ \begin{array}{l} D_1^+ \\ D_2^+ \\ D_3^+ \end{array} \right.$	72	73	85	85	84	89	87
Abundance		23	20	13	11	12	8	8
		5	7	2	4	4	3	5

of the tail on the low energy side of the experimental curve increases with pressure in the accelerator and spectrometer and is attributed to scattering and energy loss by collision in the high vacuum side of the apparatus.

In adapting this ion source for use on our high voltage accelerators we have found it more convenient to connect the gas leak and Pirani gauge through the base of the ion source. To increase the focusing efficiency the outside diameter of the probe has been reduced from 1/4 in. to 3/16 in. The same high atomic percentage has been measured with this arrangement. With 3 kv. extraction volts and 10 ma. probe current no difficulty has been found in focusing a $\frac{1}{2}$ ma. beam of ions in a 6 mm. diameter spot after passing through 3 metres of the accelerating tube.

The Clarendon Laboratory,
Oxford.

31st August 1948.

P. C. THONEMANN.
J. MOFFATT.
D. ROAF.
J. H. SANDERS.

BAYLEY, A. J., and WARD, A. G., 1948, *Canad. J. Res.*, **26**, 69.

HERTZBERG, G., 1927, *Ann. Phys., Lpz.*, (4), **84**, 565.

THONEMANN, P. C., 1946, *Nature, Lond.*, **158**, 61.

The Absorption Coefficient of the β Rays from $^{35}_{16}\text{S}$ in Aluminium

During the course of some measurements using radio sulphur $^{35}_{16}\text{S}$ it became necessary to make an estimate of the correction due to self absorption of the β rays. The end point of the β -ray spectrum is known (Libby and Lee) but as this lies outside the range of Feather's empirical formula it is hoped that the following experimental determination of the absorption coefficient in aluminium may be of use to other workers using counters with cylindrical symmetry.

The measurements were made with a demountable counter; it differs only in details from that used by Collie and Morgan in 1939. The active barium sulphate was distributed as evenly as possible as a film whose mean weight was less than 1/10 mg/cm² on the inside

of a movable glass cylinder. The counter was filled with a carefully controlled air alcohol mixture using a bellows manometer of the type described by East and Kuhn (1946). A long series of measurements showed that variations in the observed activity of a single specimen due to refilling the counter and replacing the specimen did not amount to more than 3% (including the statistical error). The aluminium foil used as an absorber was wrapped round the perforated cathode of the counter and unwrapped layer by layer. On examining the pulses with a cathode-ray tube it was observed that the aluminium foil caused the production of a small proportion of double counts. These were insufficient to interfere with the stability of the counter at the highest rate of counting used.

The absorption curve was calculated on the assumption that the number of double counts was a constant proportion of the true activity. The correctness of this assumption was verified experimentally by varying the resolving time of the counting circuit and comparing the results with those obtained with a counter free from doubles. The absorption was found to be exponential with an absorption coefficient $\mu = 780 \pm 25 \text{ cm}^{-1}$. The greatest uncertainty of the determination is due to the difficulty of measuring the exact weight of foil through which the β particles have passed. Using this figure for μ , the activity of a bare source will be halved by 2.4 mg/cm^2 of aluminium. This result may be compared with the observation of Hendricks and his collaborators (1943) that the halving mass in mica is 3 mg/cm^2 . The small difference between these results is not more than would be expected when the very different geometrical arrangements are taken into account.

We have to thank Professor Peters for putting the radio sulphur at our disposal, the Board of Education for a maintenance grant and Lord Cherwell for extending to us the facilities of his laboratory.

The Clarendon Laboratory, Oxford.

7th September 1948.

(First communicated 3rd August 1946).

C. H. COLLIE.

M. A. GRACE.

COLLIE, C. H., and MORGAN, F., 1939, *J. Sci. Instrum.*, **16**, 369.

FEATHER, N., 1938, *Proc. Camb. Phil. Soc.*, **34**, 599.

EAST, H. G., and KUHN, H., 1946, *J. Sci. Instrum.*, **23**, 185.

HENDRICKS, R. H., BRYNER, L. C., THOMAS, M. D., and IVIE, T. O., 1943, *J. Phys. Chem.*, **47**, 469.

LIBBY, W. F., and LEE, D. D., 1939, *Phys. Rev.*, **55**, 245.

The Effect of Room Temperature Radiation on the Infra-Red Response of Lead Telluride Photoconductors

One of the principal objects in studying the photoconductivity of semiconductors is to search for substances which show photosensitivity to radiation as far into the infra-red as possible. The practical importance of these substances is in their ready application to infra-red detection, and in particular to detection for the purposes of infra-red spectroscopy. It has been generally recognized that as the long wavelength limit of photoconductivity is extended, the temperature at which a detector may profitably be used must be decreased in order to cut down thermal excitation in the material. However, if the thermal radiation from the surroundings of a photoconductor contains appreciable energy at wavelengths to which the detector is photosensitive, there will be an effective background of illumination which will increase the conductivity even when no source of illumination is deliberately applied. This will be true at whatever temperature the photoconductor is maintained, unless it is surrounded by an enclosure at the same, or lower, temperature.

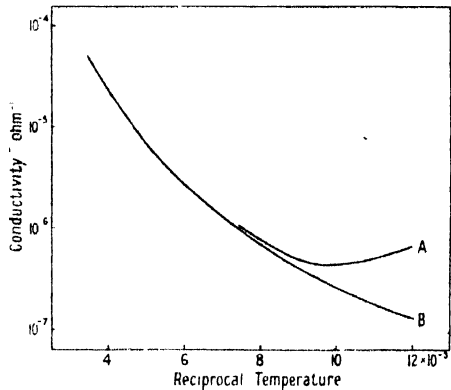
Lead telluride, in the form of thin films, has been found to show photoconductivity at least as far into the infra-red as 5.5μ , and it has been shown that at -190°C . its performance as an infra-red detector surpasses that of a good Hilger-Schwartz thermocouple for wavelengths shorter than 4.7μ (Simpson, Sutherland and Blackwell 1948). It has recently been discovered that the conductivities of all photoconductive lead telluride films are much less when they are cooled to -190°C . and surrounded by an enclosure at that temperature than when they are cooled to -190°C . but surrounded by an enclosure at room temperature. That this effect is due to the thermal radiation which in the latter

case is emitted by the enclosure at 15°C . is demonstrated in the figure where the logarithm of the conductivity of a film is plotted against the reciprocal of the temperature. Curve A shows the conductivity that is observed when the film is allowed to warm up from -190°C . to room temperature, the photoconductor being exposed all the time to an enclosure at 15°C . It will be observed that below -170°C . ($1/T > 9.7 \times 10^{-3}$), the conductivity decreases with increasing temperature; this in itself is anomalous for a semiconductor. The anomaly is removed in curve B where the conductivity of the same film is shown under conditions in which the thermal radiation was excluded by surrounding the film by an enclosure at -190°C . The conductivity at -190°C . was more than six times greater when the film was exposed to the room temperature radiation; this effect decreased rapidly as the temperature was increased, and was negligible at temperatures above -140°C .

The effect of the thermal radiation on the photoconductive response may best be illustrated by considering the following equation which is found experimentally:

$$(\sigma_1/\sigma_d)^2 = 1 + bI,$$

where σ_d is the conductivity at -190°C . with the thermal radiation excluded, and σ_1 the conductivity under an intensity of illumination I watts/cm². The constant b depends upon the wavelength of the incident light and was $2.2 \times 10^6 \text{ watt}^{-1} \text{ cm}^2$ at 2.2μ for the film shown in the figure. Using this equation, the voltage change across a film caused by the absorption of radiation while the film is passing a fixed current may be readily calculated. It is found that for an exciting current of $1 \mu\text{A}$. the voltage output is $9.2 \text{ V}/\mu\text{W}/\text{cm}^2$, $\lambda = 2.2 \mu$, when the film is entirely shielded from thermal radiation. The output for the same exciting current when the thermal radiation is not excluded is only $0.053 \text{ V}/\mu\text{W}/\text{cm}^2$.



In order to assess the effect of the thermal radiation on the ultimate limit of sensitivity of lead telluride further data are required concerning the noise characteristics of the films.

Clark Jones (1947) has shown that fluctuations in the radiation field set the ultimate limit for any detector operating under prescribed conditions. It has been shown here that by excluding the background of thermal radiation the signal output may be increased; the signal-to-noise ratio will be further increased by this means if the noise output of the film is predominantly due to statistical fluctuations in the background radiation. There is at present insufficient evidence for it to be certain that this is in fact the case.

The spectral response curve for lead telluride (Simpson, Sutherland and Blackwell 1948) has so far only been obtained with films exposed to room temperature radiation; there is a possibility that under these conditions the photosensitivity at long wavelengths is at least partly saturated. If this were so the photoconductive response would be increased even more at long wavelengths than at the wavelength 2.2μ which has been used as standard in this discussion. It seems possible that the changes in the spectral response of PbS, PbSe and PbTe obtained by Moss (1948) by operating at very low temperatures are not unconnected with the above results.

I am indebted to Dr. G. B. B. M. Sutherland and Dr. E. F. Daly for helpful discussions.

O. SIMPSON.

Cavendish Laboratory, Cambridge.

12th August 1948.

JONES, R. C., 1947, *J. Opt. Soc. Amer.*, **37**, 879.

MOSS, T. S., 1948, *Nature, Lond.*, **161**, 766.

SIMPSON, O., SUTHERLAND, G. B. B. M., and BLACKWELL, D. E., 1948, *Nature, Lond.*, **161**, 281.

REVIEWS OF BOOKS

Radar Beacons, edited by ARTHUR ROBERTS. Pp. xx+489 with 246 figures.
(New York: McGraw-Hill Book Co. Inc., 1947.) \$6.

After world war II the Radiation Laboratory of the Massachusetts Institute of Technology undertook the heavy task of preparing for publication in book form the war-time developments of Radar and Radar Techniques. Several volumes of this series have already appeared and are proving of real value to the specialist and student alike. The present volume is Volume III of the series and deals with the important subject of Radar Beacons.

A radar beacon is a small receiver transmitter apparatus which is used at the target to be located. It is actuated (or triggered, to use the correct technical term) by the pulsed signals sent out by the radar set, and in turn the beacon sends out pulsed signals which may be received at the radar set. The time delays in the beacon system can usually be made very small and constant so that range finding by timing the interval between the radar transmitted pulse and the pulse echo received back from the beacon is possible as for "primary radar". This latter depends on the target itself reflecting back some of the incident energy. The beacon system (called by Watson-Watt "a secondary radar system") has, however, several advantages when circumstances allow its use. Thus it allows target location in circumstances where ground clutter would prevent the operation of a primary radar system, greater range and, if the beacon transmissions are coded, positive identification of the target echoes. For these, and for other reasons, the beacon system was much used during the last war for general navigation of aircraft and surface vessels, for approach systems at airfields, for paratroop dropping, for bombing aids and for other purposes, and it is now finding application in civil aviation and in maritime service. The present book attempts to give a survey of these developments and, it can be said at once, that the 24 authors who have collaborated in this work have succeeded in providing a very valuable treatise which should be of special interest to all those who want to obtain a thorough understanding of the present state of the beacon art.

The book is divided into four parts: Basic Considerations, Beacon Design, Interrogator and System Design, and Beacons in the Field. The first part is especially valuable and gives a thorough treatment of the principles of the subject. After a brief introductory section dealing with the types and functions of beacons it gives a detailed discussion of the requirements of systems using beacons, considering in turn range requirements, propagation and coverage, choice of frequency, use of coding and communications, and questions of traffic and capacity. All this is explained in a very lucid manner and should be easily assimilated by the would-be beacon designer.

Part II deals with beacon design and includes sections on the design of all the component parts of the beacon. There are chapters on the design of the R.F. components, beacon receivers, interrogation coding, response coding, beacons as pulse communication systems, the design of modulators and transmitters and the associated problems of power supplies and performance testing. This is all very well explained, particularly so far as the components of microwave systems are concerned, but, in some cases, the treatment of the component parts of the longer wave systems is dealt with far too briefly.

Part III gives details of the Interrogator and system design and includes chapters on radar design for beacon operation, the design of interrogator-responders and the design of typical beacon systems. This latter chapter is well worth careful study as an outline is given of the application of the principles discussed in Part I to meet requirements of different types of beacons and will therefore be of particular value to the designer.

Part IV discusses beacons in the field and includes chapters on installation, operation and maintenance. Questions of this sort tend to be forgotten in many books and it is pleasing to a reviewer to find these subjects so well presented.

The book can be recommended with every confidence.

DENIS TAYLOR.

THE PROCEEDINGS OF THE PHYSICAL SOCIETY

VOL. 61, PART 6

1 December 1948

No. 348

The Measurement of the Eye Movements of a Prone Observer during Monocular Fixation

By MARY P. LORD

Imperial College, London

MS. received 2 January 1948, in amended form 26 March 1948

Read at a meeting of the Optical Group 26 November 1947

ABSTRACT. A critical survey of previous work on the measurement of fixation eye movements shows that none of it has satisfied all the conditions : (i) possibility of detection of movements of magnitude one minute of arc; (ii) little interference with the natural state of the eye; (iii) satisfactory treatment of the head movement problem.

A technique which more nearly meets these requirements is described. Photoelectric recording of the movements of an ultraviolet beam reflected from the surface of the cornea enables the first and second conditions to be satisfied. The reflected beam is divided into two parts, one of which falls on a horizontal straight edge, the other on a vertical straight edge. In each case more or less of the radiation passes the straight edge as the eye moves, and is focused on an electron multiplier phototube. The output of each multiplier is amplified and fed to a cathode-ray oscillograph, the time-base of which is suppressed. The oscillograph beams are thus arranged to give vertical traces only, and these are photographed simultaneously on a continuously moving film travelling in the horizontal direction. The eye movements can be deduced from the two records on the film. In an attempt to meet the third requirement the subject is placed in the prone position, in which the movements of the ultraviolet beam due to head movements are smaller than for other positions of the subject.

No attempt at analysis of the observations has yet been made.

§ 1. INTRODUCTION

THE measurement of fixation eye movements is of importance because an adequate theory of many visual phenomena must take into account any movements of the retinal mosaic over the image formed by the eye during fixation. Fixation eye movements have been studied by many previous workers, but the methods leave much to be desired ; the treatment of the head movement problem is, in general, inadequate and high sensitivity has been obtained at the expense of introducing foreign bodies on to the surface of the eye.

§ 2. EXPERIMENTAL

The requirements of an improved technique were taken as : (i) possibility of detection of movements of magnitude one minute of arc ; (ii) little interference with the natural condition of the eye ; (iii) satisfactory treatment of the head movement problem.

In a preliminary investigation, the deflection of a beam of light reflected from a plane mirror attached to a spectacle frame worn by the subject was used as an indication of the head movements. It was found that when the subject was sitting up, whatever the support used for the head, the movements were about ten times greater in magnitude than those of the subject in the prone position

biting on a dental impression. This position was therefore adopted. The subject's head rested on a firm foundation of variable height and was partially supported in a sling (figure 1). For convenience, most other parts of the equipment, the lay-out of which is indicated in figure 2, were mounted on the floor.

The principle of the system is to direct a beam of ultraviolet light ($\lambda = 3650 \text{ \AA}$) on to the cornea C in such a direction that the fraction of the light transmitted through the cornea falls on the subject's blind spot when some test object is fixated on the fovea with the illuminated eye: the subject is then, in effect, only aware of the fixation target. Part of the radiation is, however, reflected at the cornea; this reflected beam strikes a half-aluminized mirror (M_1) and is there divided so that one part falls on a horizontal straight edge (E_H) and the other on a vertical one (E_V). In each case more or less radiation passes the straight edge as the eye moves; the fraction passed is focused on an electron multiplier photocell (EM_H for the transmitted, EM_V for the reflected beam). The output of each multiplier is amplified and fed to a cathode-ray oscillograph (with blue screen), the time-base of which is suppressed. The two oscillograph beams are in this way arranged to give vertical traces only which are photographed simultaneously on a continuously moving film travelling in the horizontal direction. The eye movements can be deduced from the two records on the film.

The various parts of the apparatus will now be described in detail.

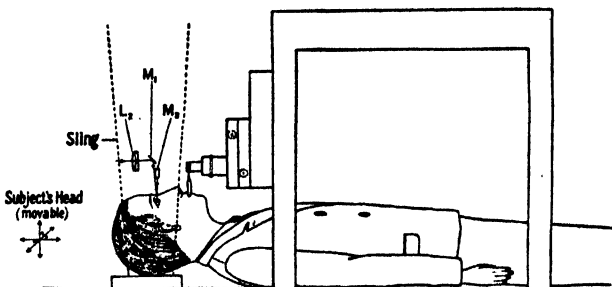


Figure 1. The subject in position.

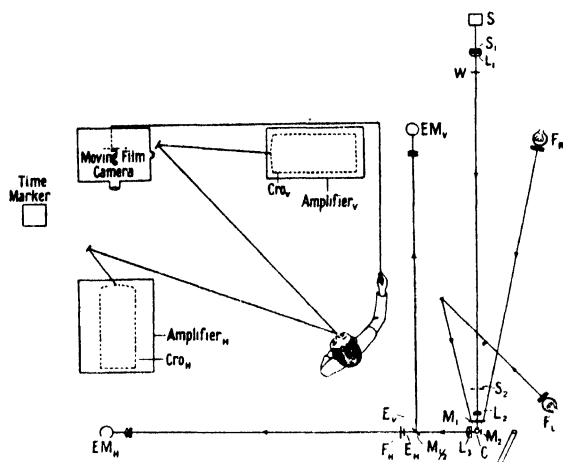


Figure 2. The lay-out of the apparatus.

(i) The optical system

The corneal illumination is obtained from a 250-watt high pressure mercury vapour lamp S operating from car batteries. S is imaged by the lens L_1 (situated immediately in front of the square aperture S_1 of side 2 mm.) in the plane of the square aperture S_2 of side 8 mm.; S_2 is situated at the focus of the lens L_2 . After passing through the lens L_2 , the light is reflected at the subject's cornea by the plane aluminized mirror M_1 inclined at approximately 45° to the vertical. The incident radiation was largely confined to $\lambda 3650 \text{ \AA}$. by the Wood's glass

filter W ; the maximum value of its energy is about $3 \mu W$. The part of the ultra-violet beam reflected at the cornea strikes the plane steel mirror M_2 (inclined at about 45° to the vertical and at 90° to M_1) centrally, and is there reflected along the axis of L_3 to M_1 . The angle of incidence at the cornea is about 3° . The illumination in the plane of the straight edges is sensibly uniform since they are imaged on S_1 by the lens L_3 , the subject's cornea and the lens L_2 . The position occupied by the beam under these conditions is marked by spots of phosphorescent paint on a fluorescent screen (F_H) which can be swung into position immediately behind E_H when required.

R.C.A. 931 A. multipliers were used since, as far as is known, they are the most sensitive available; the tubes employed were specially selected for minimum fatigue. From knowledge of the sensitivity distribution of the photo-cathode, the most suitable lens systems for condensing the ultraviolet beam on the multipliers were calculated and the best from those available selected by trial: a sphero-cylindrical lens giving an image 2 mm. long and $\frac{1}{4}$ mm. wide on the multiplier cathode was used in each case. The multiplier response was then

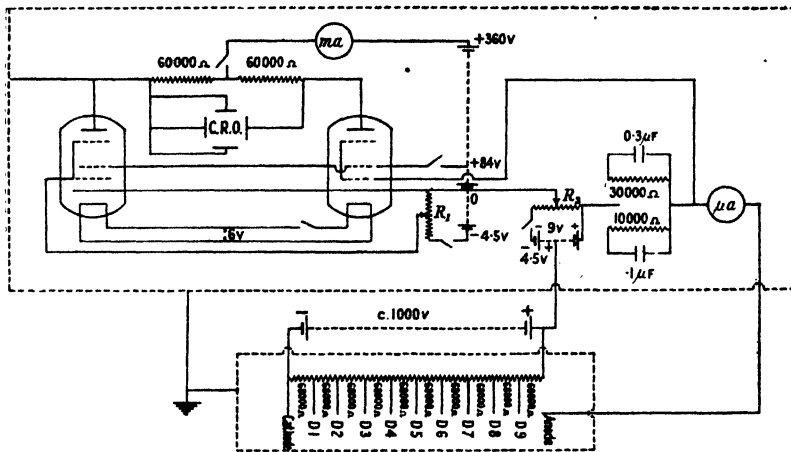


Figure 3. Circuits used with each multiplier.

sensibly uniform for positions of the cornea in a 5 mm. range around that for which S_1 is imaged on the straight edges; the size of the patch on the multiplier was, moreover, independent of the position of the straight edges.

The system was calibrated by using a lens with the same curvature as that of the cornea. The straight edges were set so that each cut off approximately half the patch when this was central. The responses of EM_H and EM_V were then recorded for complete movement of the beam over the straight edges for each of a series of vertical positions of the "artificial cornea" 1 mm. apart and lying in the 5 mm. range surrounding the position for which S_1 is imaged on the straight edges. Apart from the initial and end portions all the curves were linear. The sensitivity reaches a maximum value ($3.7 \mu A/\text{min.}$) for the position in which S_1 is imaged on the straight edges and falls to about $3 \mu A/\text{min.}$ at either end of the 5 mm. range.

The mirror M_1 also served to reflect the fixation beam into the eye. To enable observations to be made on either eye, two fixation targets were available. Each was illuminated by a filament lamp: F_R and F_L for the right and left eyes

respectively. For subjects with refractive errors, correcting lenses were inserted in the path of each beam; neutral and coloured filters could also be inserted.

(ii) *The electrical system*

The lay-out of the circuits used in conjunction with either multiplier is shown in figure 3; each circuit is mounted in two earthed tin boxes as indicated by the dotted lines. The signal fed to the amplifier is obtained by passing the multiplier output through a resistance (either 10 or $30 \times 10^3 \Omega$). The purpose of the condenser connected across this resistance is to suppress the major portion of the background noise from the multiplier; the cut-off frequency of the amplifier is then about 1000 c/sec. There is therefore little attenuation of frequencies of 100 c/sec.—the highest found in eye movements by other observers.

The effect of fluctuations in the heater voltage and H.T. supplies are largely eliminated by the use of H.F. pentodes in the single-stage balanced D.C. amplifier. The rheostat R_1 is permanently set so that the compensating valve works at the centre of its operating range. With the spot at the centre of the oscillograph screen, the positions of R_2 were marked for various microammeter readings, for low and high gain; change from one position to another could then easily be made. The amplifiers were found to have good stability: in figure 4 (a) is a record of the performance for no signal from the multipliers. The amplifiers are, moreover, sensibly linear; from the relationship between microammeter reading and screen deflection, it appears that the maximum value of the sensitivity in either direction is about 6 mm. screen deflection per minute; deflections due to noise from the multipliers do not, in general, correspond to more than $\frac{1}{2}$ minute of arc; a record from the stationary artificial cornea is shown in figure 4 (b).

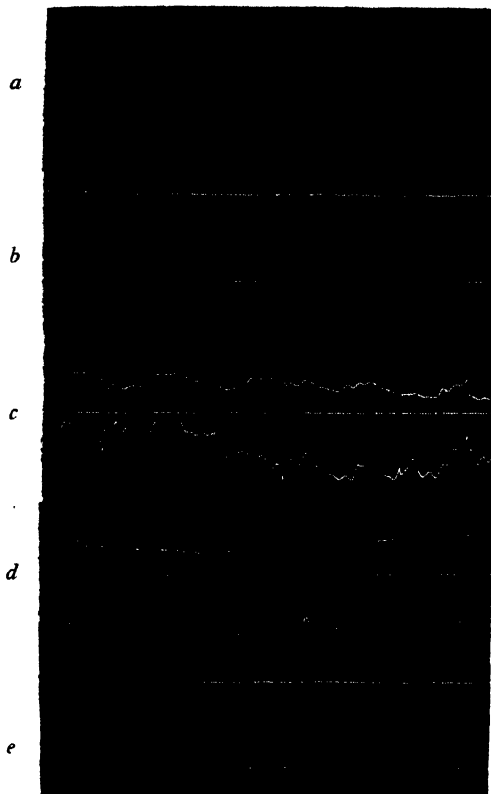


Figure 4.

- (a) Amplifier performance for no signal from multipliers.
- (b) Record from stationary artificial cornea.
- (c) Eye movements during fixation.
- (d) Head movements.
- (e) Record from an eye in the absence of the straight-edges.

Time marked by dots, 50/sec.; amplitude shown at side: length of line between arrow heads represents 10 minutes of arc.

(iii) *Positioning the subject*

As a first approximation, the rod R is swung over the eye under investigation : when the head height is correct, the cork-covered peg attached to R just touches the eyelid. After a few minutes in the dark, the subject is asked to bite on the dental impression and look for a blue haze—a fluorescence effect in the eye due to the ultraviolet irradiation. To facilitate the fine adjustment of the height of the head, the square S_2 is diaphragmed into an L and the head height adjusted until the haze appears square and as small as possible : when the head is too high, the haze appears as an upright L ; when it is too low, as an inverted L. The beam is observed on F_H and the subject's head moved in the horizontal plane until the patch is central. The subject is then asked to observe the position of the blue haze when looked at directly relative to the fixation target and the blue and red images of S_2 . The straight edges are then set in the appropriate position as marked by spots of phosphorescent paint ; movements of the eye from side to side produce changes in the amount of light passing E_H ; those up and down the length of the observer cause changes in the amount passing E_V .

§ 3. RESULTS

The results so far have been confined to the investigation of the possibilities of the method. It appears that eye movements of appreciable magnitude may sometimes be present. Figure 4 (c) shows a portion of the record of an observer for fixation of the centre point of a cross, angular length of arm $20'$, width $1'$. Figure 4 (d) shows the effect of head movements on those of the corneal reflex obtained by supporting the artificial cornea on a spectacle frame immediately above the live one. By calculation, the corresponding compensatory eye movements are negligible. In figure 4 (e), the record obtained from a live cornea in the absence of the straight edges is shown.

§ 4. DISCUSSION

It appears from the preliminary records so far obtained, that the method satisfies the three requirements laid down at the beginning of the experimental section more closely than any previously used. Simultaneous recording of the artificial corneal reflex would be a desirable feature and it is hoped to make arrangements for this. It may be argued that most visual characteristics involving eye movements have been investigated under conditions different from those described in this paper and that such measurements might be irrelevant to those problems. Whether or not this is a valid criticism will appear in the course of the investigations since it is anticipated that these will include the study of the various characteristics under the conditions described.

ACKNOWLEDGMENTS

The author is very grateful to Dr. W. D. Wright for suggesting the problem, for many helpful suggestions and criticisms, and for acting as subject or operator. She is glad to acknowledge the help given by many members of the Physics Department Staff in the construction and assembling of the apparatus. Thanks are also due to the subjects, especially to Mr. R. W. G. Hunt. The author gratefully acknowledges the financial support of the Medical Research Council.

On the Designing of Aspheric Surfaces

By E. WOLF

H. H. Wills Physical Laboratory, University of Bristol

Communicated by E. H. Linfoot; MS. received 6 December 1947, in amended form 14 April 1948

ABSTRACT. A method is developed for the design of an aspheric surface to ensure axial stigmatism in any given centred optical system. It is shown that when the surface is in the interior of the system the determination of its profile depends on the solution of a certain transcendental equation. An iterative method is described which gives the profile to any desired degree of accuracy. When the corrector is the last surface of the system the solution is greatly simplified and for such cases exact parametric equations, involving quantities which can easily be obtained from a ray trace, are given.

§ 1. INTRODUCTION

THE excellent performance of the Schmidt camera, as well as the increasing demand for improved image qualities of optical systems, has led in recent years to the construction of a number of systems which employ aspheric surfaces. This requires the development of methods which allow the designer to determine what shapes such surfaces are to take in order to obtain good performance over a finite field. With the exception of the elegant "see-saw" diagram of Burch (1942) no method is at present available for dealing with this question. Burch's method is restricted to the extended paraxial region, i.e. the region in which aberrations above the third order (i.e. aberrations higher than Seidel's) can be neglected. This method enables us to find the most favourable range of the design parameters and is particularly useful in the early stages of the design. In systems where higher aberrations have to be taken into account the Seidel design needs to be supplemented by a further stage of approximation, involving the modification of one or more surfaces of the system.* Although in general one aims at the optimum performance over the whole of a finite field, it is often sufficient to design the correcting surface in such a way that a particular aberration is corrected at one point of the field.

The present paper is concerned with the design of a correcting surface ensuring axial stigmatism in any given centred system. Methods which deal with this question have been previously described by Linfoot (1943), Luneberg (1944), Herzberger and Hoadley (1946) and Wolf and Preddy (1947).

The method of Linfoot is suitable for the design of two aspheric surfaces in certain astrographic cameras, ensuring both the absence of axial spherical aberration and the satisfaction of the exact sine condition. The method is still applicable if we aim only at the design of a correcting plate to attain axial stigmatism in systems where, as in the Schmidt and Meniscus-Schmidt cameras and Schmidt-Cassegrain systems, the corrector is the first or last surface.

Whenever the aspheric surface is adjacent to one of the foci its profile may be determined by the exact equations of Luneberg (related to equation (4.6) of the present paper) which he obtained by the use of characteristic functions.

* This problem is examined by D. S. Volosov (*J. Opt. Soc. Amer.*, 1947, **37**, 342).

A method with a wider range of application was developed by Herzberger and Hoadley. It can be used for the design of a correcting surface which is situated anywhere in the system. This method is based on the computation of a number of approximate surfaces, which approach the ideal.

More recently, Preddy and the writer have shown how to design a plano-aspheric lens which would give axial stigmatism to a given symmetrical pencil of rays. The present paper gives an extension of our analysis which applies to the design of a correcting surface, whether or not it is situated in the interior of the system. The computation is based on exact equations for the surface in question. These equations involve quantities which can be easily obtained from a ray trace. In some cases the use of iterative methods is required, but the accuracy reached at every stage may be estimated without the additional ray traces required in the method of Herzberger and Hoadley. When the corrector is adjacent to one of the foci no iteration is needed, and for such cases exact parametric equations are given.

§ 2. PRELIMINARY RESULTS

The solution of our main problem will be based on some properties of normal rectilinear congruences (N.R.C.) which we discuss in this section. For the purpose of dealing with axially symmetrical systems, it is sufficient to consider the case of two dimensional N.R.C.s.

(i) Let Γ be a plane N.R.C. of rays. We choose rectangular axes OX , OY and denote by h the height at which a typical member of Γ meets OY , or would meet OY if produced sufficiently far. Further, we denote by ω ($-\pi < \omega \leq \pi$) the angle which this typical ray makes with OX ; we measure ω anticlockwise from the positive x direction to the direction along which the light advances. The congruence is then completely specified by a functional relation between ω and h say

$$\omega = \omega(h) \quad \dots\dots(2.1)$$

where h is regarded as a free parameter. We shall refer to a ray, determined by the value h_r of the parameter h , as the ray h_r and denote by ω_r the corresponding value of ω . The following simple theorem can then be proved * by differential geometry or otherwise :

Theorem A : If two rays h_1 and h_2 intersect a wave front W of Γ in P_1 and P_2 , and OY in Q_1 and Q_2 (figure 1), then the optical path difference $[P_2Q_2] - [P_1Q_1]$ is given by

$$[P_2Q_2] - [P_1Q_1] = n \int_{h_1}^{h_2} \sin \omega \, dh \quad \dots\dots(2.2)$$

where n is the refractive index of the medium in which the rays are situated.

(ii) With the help of theorem A we now examine the (1, 1) correspondence between the parameters of two N.R.C.s which are transformed into each other by refraction or reflection. In general the possibility of such a transformation is asserted by the converse of Malus' theorem (Levi-Civita 1900).

Let Γ and Γ' be two coplanar N.R.C.s, situated in spaces of respective refractive indices n and n' , and let Σ be a refracting (or reflecting) profile which transforms these two congruences into each other. To simplify the analysis

* See for example Wolf and Preddy (1947), p. 705. It was assumed there that the congruence is symmetrical and $n=1$; but it can easily be verified that these restrictions are not necessary. The difference in signs of the integrals is due to a different sign convention used in the two papers.

we refer Γ and Γ' to the same Cartesian axes with their origin coincident with a point of Σ . Let h_r and h'_r be any two corresponding rays in this transformation; that is to say, the ray h_r of Γ is transformed by refraction at Σ into the ray h'_r of Γ' (figure 2). Further, let W and W' denote wave fronts of Γ and Γ' respectively. Also, let P_r and Q_r be the points of intersection of the ray h_r with W and OY respectively, and similarly for the dashed letters. (By definition $OQ_r = h_r$, $OQ'_r = h'_r$). Finally let (x_r, y_r) denote the coordinates of the point R_r of Σ where these rays meet.

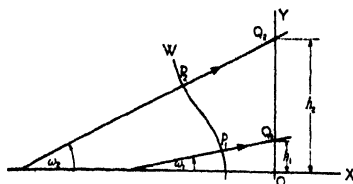


Figure 1.

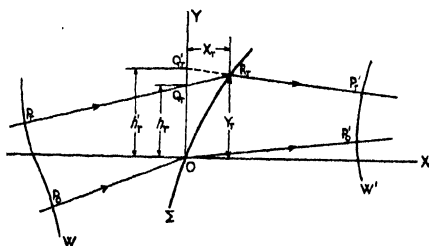


Figure 2.

Since the optical path between the two wave fronts is constant when taken along corresponding rays,

$$[P_r R_r] + [R_r P'_r] = [P_0 O] + [OP'_0], \quad \dots (2.3)$$

where the square brackets denote optical lengths and the suffix zero refers to rays which pass through O .

We have from figure 2 and theorem A, that

$$[P_r R_r] = [P_r Q_r] + [Q_r R_r] = [P_0 O] + n \int_0^{h_r} \sin \omega \, dh + n x_r \sec \omega_r \quad \dots (2.4)$$

and similarly,

$$[R_r P'_r] = -[P'_0 O] - n' \int_0^{h'_r} \sin \omega' \, dh' - n' x_r \sec \omega'_r. \quad \dots (2.5)$$

Substituting from (2.4) and (2.5) into (2.3), we obtain

$$x_r (n \sec \omega_r - n' \sec \omega'_r) + n \int_0^{h_r} \sin \omega \, dh - n' \int_0^{h'_r} \sin \omega' \, dh' = 0 \quad \dots (2.6)$$

We also have from figure 2

$$y_r = h_r + x_r \tan \omega_r \quad \dots (2.7)$$

and

$$y_r = h'_r + x_r \tan \omega'_r. \quad \dots (2.8)$$

On eliminating x_r and y_r between (2.6), (2.7) and (2.8) we obtain the following interesting invariant property :

Theorem B : *In the transformation $\Gamma \rightarrow \Gamma'$, the parameters of corresponding rays, when referred to any set of rectangular axes with their origin placed at a point of the transforming profile, satisfy the relation*

$$\sin (\omega_r - \omega'_r) \left(n \int_0^{h_r} \sin \omega \, dh - n' \int_0^{h'_r} \sin \omega' \, dh' \right) + (n \cos \omega'_r - n' \cos \omega_r) (h'_r - h_r) = 0 \quad \dots (2.9)$$

§ 3. THE DESIGN OF A CORRECTING SURFACE ACHIEVING AXIAL STIGMATISM

(i) In a number of optical systems employing an aspherical surface, the latter is designed to ensure axial stigmatism. Examples of this are the Schmidt television projector system, certain modern astrographic cameras and the semi-aplanatic reflecting microscope. With the help of the results of the previous section, we shall now show how to determine the profile of such a correcting surface for any given centred optical system.

We introduce a rectangular reference system with origin O at the pole (assumed to be known) of the correcting surface Σ and with OX along the axis of symmetry. Rays starting from the axial object point S are refracted or reflected at the intermediate surfaces and form an N.R.C. Γ in the space preceding Σ . In figure 3

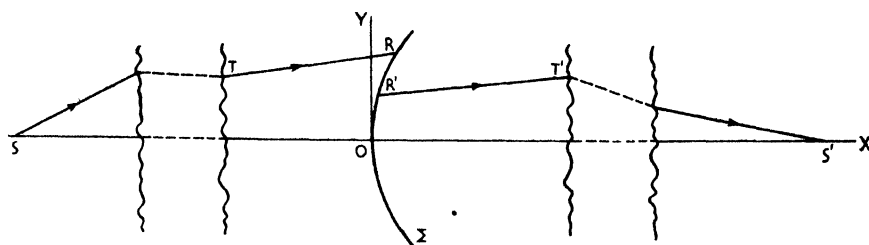


Figure 3.

a typical member of this congruence is denoted by TR . Let

$$\omega = \omega(h) \quad \dots\dots(3.1)$$

be the equation of this congruence. In practice, it is often more convenient to use instead of (3.1) a polynomial approximation of the type

$$\sin \omega = a_1 h + a_3 h^3 + \dots\dots a_{2k+1} h^{2k+1}. \quad \dots\dots(3.2)$$

This is obtained by tracing a number of rays from the axial object point to the space which precedes Σ , computing for each ray the quantities ω and h and then fitting a polynomial in the usual manner. The use of such an approximation, however, is not essential, and when great accuracy is required may not even be desirable. In such cases it may be better to use a table of values of ω with corresponding values of h based on the tracing of a thick fan of rays.

In the space which follows Σ , the rays which are traced backwards* from the ideal axial image point S' , will form another N.R.C. Let Γ' be the congruence obtained from this "backtraced" congruence, by reversing the direction of every ray, i.e. we regard the rays of Γ' as proceeding from the correcting surface towards the image point. For Γ' we again obtain a relation of the type (3.1) or (3.2). We shall use dashed letters to refer to all quantities related to Γ' .

It is clear that the required surface will transform Γ into Γ' . Hence it follows from theorem B that the parameters of corresponding rays in this transformation satisfy the relation

$$\sin(\omega_r - \omega'_r) \left(n \int_0^{h_r} \sin \omega \, dh - n' \int_0^{h'_r} \sin \omega' \, dh' \right) + (n \cos \omega'_r - n' \cos \omega_r)(h'_r - h_r) = 0. \quad \dots\dots(3.3)$$

* When the correcting surface is the last (or first) surface of the system, the backwards ray trace becomes superfluous. The simplified analysis for such cases is given in § 4.

This relation enables us to compute the profile of Σ . For if h_r is any particular ray of the first congruence, the corresponding h'_r can be obtained by solving (3.3) as a transcendental equation (§ 3 (ii)); the coordinates of the point (x_r, y_r) on Σ are then given by the equations (2.6) and (2.7), which can be written as

$$x_r + iy_r = \frac{\cos \omega'_r \exp(i\omega_r)}{n \cos \omega'_r - n' \cos \omega_r} \left(n' \int_0^{h'_r} \sin \omega' dh' - n \int_0^{h_r} \sin \omega dh \right) + ih_r. \quad \dots\dots(3.4)$$

Although it may be simpler to evaluate x_r and y_r by the use of some other combination of the relations (2.6), (2.7) and (2.8), it is better to use the above form; for any error arising in an approximate solution of (3.3) will cause as a rule smaller errors in the coordinates when these are calculated from (3.4) than in the other cases. This is easily verified by examining the Taylor's expansions of the various combinations.

In this manner we obtain the profile of the correcting surface in tabulated form. If desired this can then be approximated by a polynomial of the form

$$x = A_2 y^2 + A_4 y^4 + \dots + A_{2m} y^{2m} \quad \dots\dots(3.5)$$

obtained by curve fitting in the usual manner.

(ii) *Solution of (3.3).* We observe that (3.3) can be written in the form

$$h'_r = \psi(h_r, h'_r) \quad \dots\dots(3.6)$$

where

$$\psi(h_r, h'_s) = h_r + \frac{\sin(\omega_r - \omega'_s)}{n \cos \omega'_s - n' \cos \omega_r} \left(n' \int_0^{h'_s} \sin \omega' dh' - n \int_0^{h_r} \sin \omega dh \right) \quad \dots\dots(3.7)$$

and ω_r and ω'_s are connected with h_r and h'_s respectively by relations of the type (3.1) or (3.2).

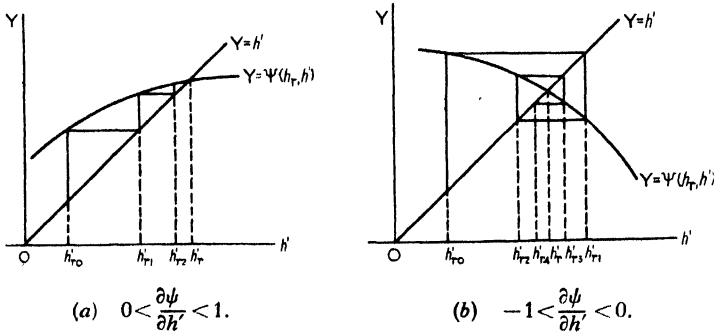


Figure 4. Illustrating the convergence of sequence (3.8).

Suppose we take an approximation (say $h'_{r,0}$) to h'_r given h_r . We then form a sequence

$$h'_{r,0}, h'_{r,1}, h'_{r,2} \dots\dots \quad \dots\dots(3.8)$$

where

$$h'_{r,i+1} = \psi(h_r, h'_{r,i}). \quad \dots\dots(3.9)$$

Such a sequence will converge to the desired root h'_r of (3.3) provided that, in the neighbourhood of this root,

$$|\partial \psi / \partial h'_s| < 1. \quad \dots\dots(3.10)$$

The geometrical interpretation of the successive values of (3.8) is shown in figures 4 (a) and 4 (b); it is seen that the rapidity of convergence depends on

the magnitude of $\partial\psi/\partial h'_s$. We find on differentiating (3.7) that

$$\frac{\partial\psi}{\partial h'_s} = \left(\frac{d\omega'_s}{dh'_s} \right) \left(n' \int_0^{h'_r} \sin \omega' dh' - n \int_0^{h_r} \sin \omega dh \right) \frac{\cos \omega_r [n' \cos (\omega_r - \omega'_s) - n]}{[n \cos \omega'_s - n' \cos \omega_r]^2} + \frac{n' \sin \omega'_s \sin (\omega_r - \omega'_s)}{[n \cos \omega'_s - n' \cos \omega_r]}. \quad \dots\dots(3.11)$$

As a rule, ω , ω' , and $d\omega'/dh'$ will be small* so that (3.10) is satisfied and the convergence of the iteration is rapid.

Suppose now that we stop at a particular term (say h'_{rt}) in (3.8) and use it in (3.4) for the calculation of x_r and y_r . We can make a rough estimate of the errors Δx_r and Δy_r thus introduced. We find, on expanding (3.4) and neglecting second and higher powers, that

$$\Delta x_r = O(\omega'_{rt} \Delta h'_{rt}); \quad \Delta y_r = O(\omega_r \omega'_{rt} \Delta h'_{rt}) \quad \dots\dots(3.12)$$

where the difference $\Delta h'_{rt}$ defined by

$$\Delta h'_{rt} = h'_r - h'_{rt} \quad \dots\dots(3.13)$$

can be estimated from the behaviour of the sequence.

The above estimates show that even a rough approximation to h'_r will often give a good approximation to x_r and y_r .†

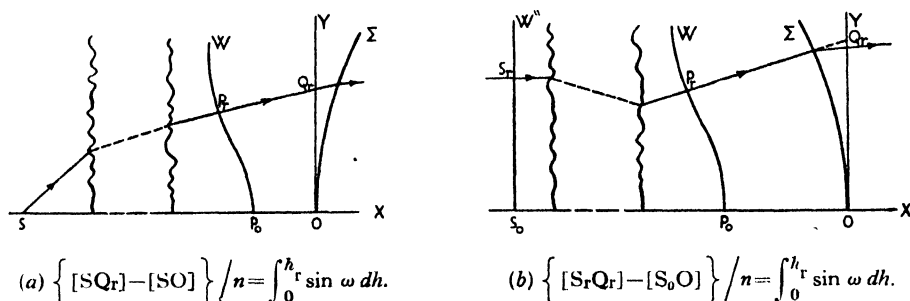


Figure 5.

The initial value $h'_{r,0}$ in (3.8) can be chosen in various ways. We can, for example, take as $h'_{r,0}$ the value of the parameter which corresponds to h_r in the congruence obtained by the refraction of Γ in the *uncorrected* surface. But even the rough approximation $h'_{r,0} = h_r$ will be adequate in most cases.

(iii) *Evaluation of $\int_0^{h_r} \sin \omega dh$.* The integral $\int_0^{h_r} \sin \omega dh$ which occurs in the previous analysis can be evaluated in several ways.

If the polynomial approximation (3.2) is used, the integral is given at once by

$$\int_0^{h_r} \sin \omega dh = \frac{a_1}{2} h_r^2 + \frac{a_3}{4} h_r^4 + \dots\dots \frac{a_{2k+1}}{2k+2} h_r^{2k+2}. \quad \dots\dots(3.14)$$

There are other methods which do not depend on an expansion for $\sin \omega$.

* The term $d\omega'_s/dh'_s$ has the following interpretation: if ρ'_s denotes the radius of curvature of the wave front of Γ which passes through the point $(0, h'_s)$ then $d\omega'_s/dh'_s = (\cos \omega'_s)/\rho'_s$.

† For example, in a Meniscus-Schmidt Projector (with the corrector facing the meniscus) of short focus 8.7 in., plate diameter 10.3 in. and throw distance 128 in., the approximation $h'_r = h_r$ gave an accuracy of 1/10 of a refraction fringe in the coordinates of the plate profile. It is generally possible thus to avoid the iteration with small loss of accuracy if the corrector is a plate.

Let S denote the axial object point. If P_r and P_0 (figure 5 (a)) lie on the same wave front W , in the space which precedes Σ we have

$$[SP_r] = [SP_0]. \quad \dots \dots (3.15)$$

From theorem A and (3.15) it follows that

$$\int_0^{h_r} \sin \omega \, dh = \frac{1}{n} \{ [SQ_r] - [SO] \}. \quad \dots \dots (3.16)$$

Similarly, if the axial object point is at infinity, we have

$$\int_0^{h_r} \sin \omega \, dh = \frac{1}{n} \{ [S_r Q_r] - [S_0 O] \}. \quad \dots \dots (3.17)$$

The meaning of S_r and S_0 is shown in figure 5 (b), where W'' is any line perpendicular to OX in the object space.

Equation (3.16) [or (3.17) when the object point is at infinity] gives the integral in terms of an optical path difference which can be obtained from ray trace. We found, however, in numerical applications that there was a large drop of significant figures when the integral was evaluated in this way, so that in such cases it may be advisable to use a larger number of figures in the computation of the paths of the rays.

If a sufficient number of rays are traced from the object space, up to the space which precedes the correcting surface, the integral may also be evaluated numerically.

The value of $\int_0^{h_r} \sin \omega' \, dh'$ can be obtained in a similar way. Owing to the iteration process an evaluation from a polynomial approximation for $\sin \omega'$ may, however, be more convenient.

§ 4. SPECIAL CASE: THE DESIGN OF A CORRECTOR WHICH IS THE FIRST OR LAST SURFACE IN A SYSTEM

(i) In practically all aspheric systems which have been constructed up to now, the corrector is the first or last surface in the system. The selection of the surface which is to be changed into a corrector depends of course primarily on the effect which the latter will have on the off-axis performance of the system. Cooke triplets of large aperture-ratio, for example, suffer badly from zonal aberration. It was thought at first that the figuring of the central concave lens, which in most designs is at or near the stop, would remove this without changing seriously the off-axis errors. Higher coma was, however, in consequence considerably aggravated, presumably because the correction increased the departure from the exact sine condition. It was found better to figure the last surface and to re-balance the coma by slightly modifying the design.

When the corrector is the last surface of the system the computation of its profile is considerably simplified. In such a case the congruence Γ' is given by

$$\tan \omega' = (\text{constant}) \cdot h' \quad \dots \dots (4.1)$$

and the equations of the corrector can be obtained from our previous analysis by eliminating ω' and h' between (3.3), (3.4) and (4.1). As the elimination is rather laborious, we shall deduce the solution directly.

Let Γ be the congruence in the space which precedes the last surface, and let S' be the ideal axial image point to which all the rays are to converge after refraction at this surface. We denote by s' the distance OS' measured along the positive direction of the axis and retain the previous symbols for the other quantities (figure 6). We drop the suffix r as no confusion is likely to occur.

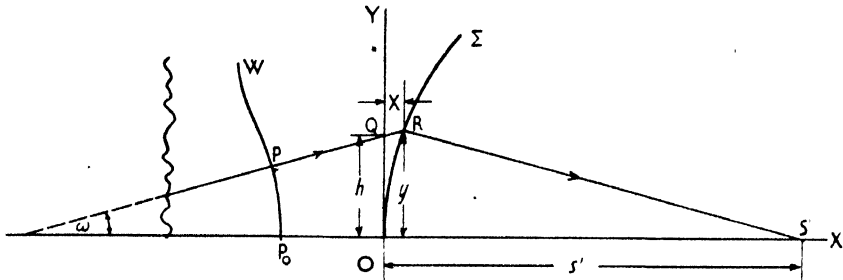


Figure 6.

Since the optical path from the wave front W to S' is constant

$$[PQ] + nx \sec \omega + n'[(s' - x)^2 + (h + x \tan \omega)^2]^{1/2} = [P_0O] + n's'. \quad (4.2)$$

From (4.2) and theorem A it follows that

$$n \int_0^h \sin \omega dh + nx \sec \omega + n'[(s' - x)^2 + (h + x \tan \omega)^2]^{1/2} - n's' = 0. \quad (4.3)$$

Rationalizing (4.3), we find that x satisfies the equation

$$Ax^2 + 2Bx \cos \omega + C \cos^2 \omega = 0 \quad (4.4)$$

where $A = n^2 - n'^2$;

$$B = n^2 \int_0^h \sin \omega dh - nn's' + n'^2 (s' \cos \omega - h \sin \omega);$$

$$C = \left(n \int_0^h \sin \omega dh \right) \left(n \int_0^h \sin \omega dh - 2n's' \right) - n'^2 h^2.$$

Solving (4.4) and using the relation

$$y = h + x \tan \omega \quad (4.5)$$

we finally obtain the following exact equation for the correcting surface in terms of the free parameter h :

$$x + iy = -(e^{i\omega}/A)[B \pm (B^2 - AC)^{1/2}] + ih. \quad (4.6)$$

Using the boundary condition $x = y = 0$ when $h = \omega = 0$, we find that the positive or negative sign of the square root is to be taken according as $n' < n$ or $n' > n$. The integral which occurs in B and C can again be evaluated by one of the methods discussed in § 3 (iii).

In order to avoid confusion about the value of ω , we emphasize that in applying (4.6) or any other equation of § 4 we regard the rays of the congruence Γ as proceeding towards the correcting surface. Thus, for instance, in the Meniscus-Schmidt

system (figures 8 (a), 8 (b)), when calculating the correcting surface, we interchange the object and image points. In figure 7, which shows a Schmidt projector suitable for cathode tube projection a typical member of the congruence Γ is denoted by TR; h is then the height at which TR produced meets OY. The corresponding value of ω lies between $-\pi/2$ and 0.

(ii) *Refracting surfaces, $s' = \infty$.* When one of the foci is at infinity or when the correcting surface is a mirror the solution simplifies still further.

Proceeding to the limit of (4.6) as $s' \rightarrow \infty$ (or setting $\omega' = 0$ for all values of h' in (3.4)), we obtain the following equation for the correcting surface in systems working with one focus at infinity :

$$x + iy = \frac{ne^{i\omega}}{n' \cos \omega - n} \int_0^h \sin \omega dh + ih. \quad \dots\dots (4.7)$$

This equation can be used, for example, in the design of the correcting plate of the Schmidt and Meniscus-Schmidt cameras and of the Schmidt-Cassegrain systems. Clearly (4.7) applies whether the aspheric or the plane side of the plate faces the incoming light. The two cases for the Meniscus-Schmidt camera of Hawkins and Linfoot (1945) are illustrated in figures 8 (a) and 8 (b). As before, a typical member of the congruence Γ (which of course is different in the two cases) is denoted by TR.

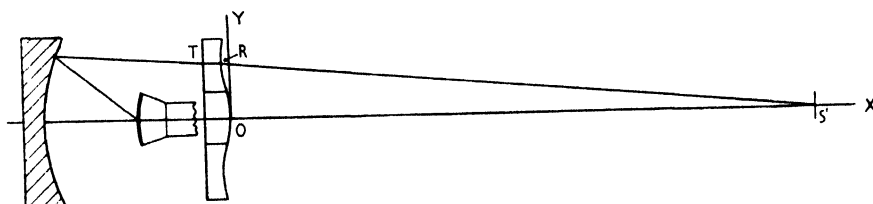


Figure 7. Schmidt projector system.

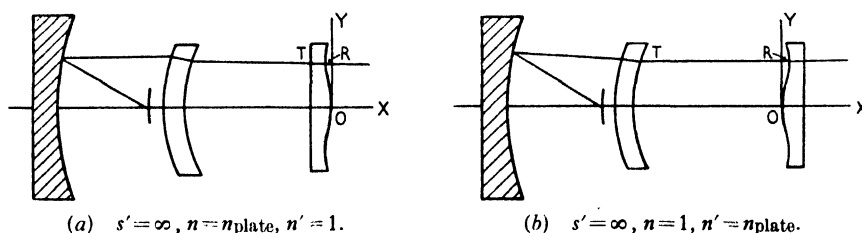


Figure 8. Meniscus-Schmidt systems.

(iii) *Reflecting surface, $s' < \infty$.* If the correcting surface is a mirror we obtain with the same sign convention and by analysis similar to that of § 4 (i) the following exact parametric equations for its profile

$$x + iy = Ce^{i\omega}/2B + ih \quad \dots\dots (4.8)$$

where

$$B = \int_0^h \sin \omega dh + s' (\cos \omega - 1) - h \sin \omega;$$

$$C = \left(\int_0^h \sin \omega dh \right) \left(2s' - \int_0^h \sin \omega dh \right) + h^2.$$

This case covers, for example, the semi-aplanat reflecting microscope recently constructed by Burch (1947) ; its optical system is shown in figure 9.

(iv) *Reflecting surface*, $s' = \infty$. Proceeding to the limit of (4.8) as $s' \rightarrow \infty$, we obtain the following equations for a correcting mirror in systems working with one focus at infinity :

$$x + iy = \frac{e^{i\omega}}{\cos \omega - 1} \int_0^h \sin \omega \, dh + ih. \quad \dots\dots (4.9)$$

This case is illustrated in figure 10 by a modified Cassegrain telescope with figuring on the concave primary mirror.

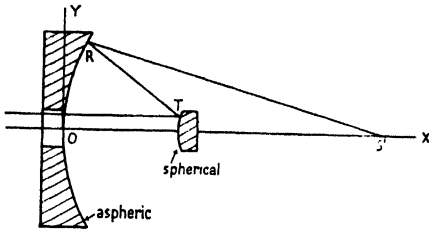


Figure 9. Semi-aplanat reflecting microscope.

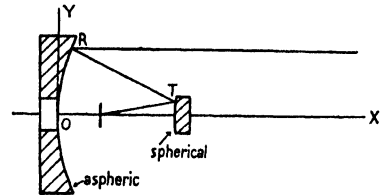


Figure 10. Modified Cassegrain telescope.

It will be noted that (4.6), (4.7), (4.8) and (4.9) are exact parametric equations of Hamilton's focal refractors and reflectors (Hamilton 1931, Synge 1937), with focus S' on the axis of an arbitrary symmetrical congruence Γ defined by the relation $\omega = \omega(h)$.

ACKNOWLEDGMENTS

The author wishes to thank Dr. E. H. Linfoot for his encouragement and valuable criticisms. He is also indebted to Mr. W. S. Preddy and Dr. G. D. Wassermann for several useful discussions.

REFERENCES

- BURCH, C. R., 1942, *Mon. Not. R. Astr. Soc.*, **102**, 159; 1947, *Proc. Phys. Soc.*, **59**, 47.
 HAMILTON, Sir W. R., 1931, *Collected Papers* (Cambridge : University Press), Vol. 1, pp. 11, 89.
 HAWKINS, D. G., and LINFOOT, E. H., 1945, *Mon. Not. R. Astr. Soc.*, **6**, 334.
 HERZBERGER, M., and HOADLEY, H. O., 1946, *J. Opt. Soc. Amer.*, **36**, 334.
 LEVI-CIVITA, T., 1900, *R. Acc. Lincei Rend.*, 237.
 LINFOOT, E. H., 1943, *Proc. Phys. Soc.*, **55**, 493; 1944, *Mon. Not. R. Astr. Soc.*, **104**, 48.
 LUNEBERG, R. K., 1944, *Mathematical Theory of Optics* (Providence, R.I. : Brown University).
 SYNGE, J. L., 1937, *Geometrical Optics* (Cambridge : University Press), p. 21.
 WOLF, E., and PREDDY, W. S., 1947, *Proc. Phys. Soc.*, **59**, 704.

The Interference Pattern of Two Neighbouring Radiations in a Hilger Prism Interferometer

By L. C. TSIEN *

Institute of Physics, National Academy of Peiping, Peiping, China.

* Now at Academia Sinica, Nanking, China

*Communicated by E. N. da C. Andrade ; MS. received 12 November 1946,
and in amended form 8 April 1948*

ABSTRACT. When two neighbouring radiations are applied to a Hilger prism interferometer, an interference pattern of two sets of fringes is obtained. These sets can be arranged at different orientations, when a system of grey bands results from their intersection. Characteristics of these bands are described, and a theoretical consideration is given. The possibility of applying the effect to the resolution of spectral lines is discussed.

§ 1. INTRODUCTION

IN the optical industry the Hilger interferometer is usually used with a Hewittic mercury vapour lamp together with a suitable filter, so that substantially monochromatic light is used corresponding to the predominant radiation 5461 Å. The interference fringes employed are similar to Newton's rings. The present paper deals both experimentally and theoretically with the effects produced when two neighbouring radiations, such as those of the sodium yellow doublet, are employed.

§ 2. EXPERIMENTAL

A Hilger prism interferometer with mirrors 85 mm. square and a large flint

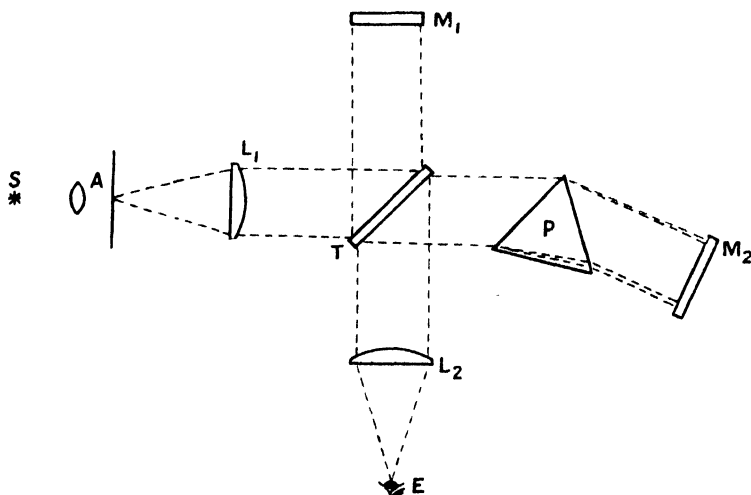
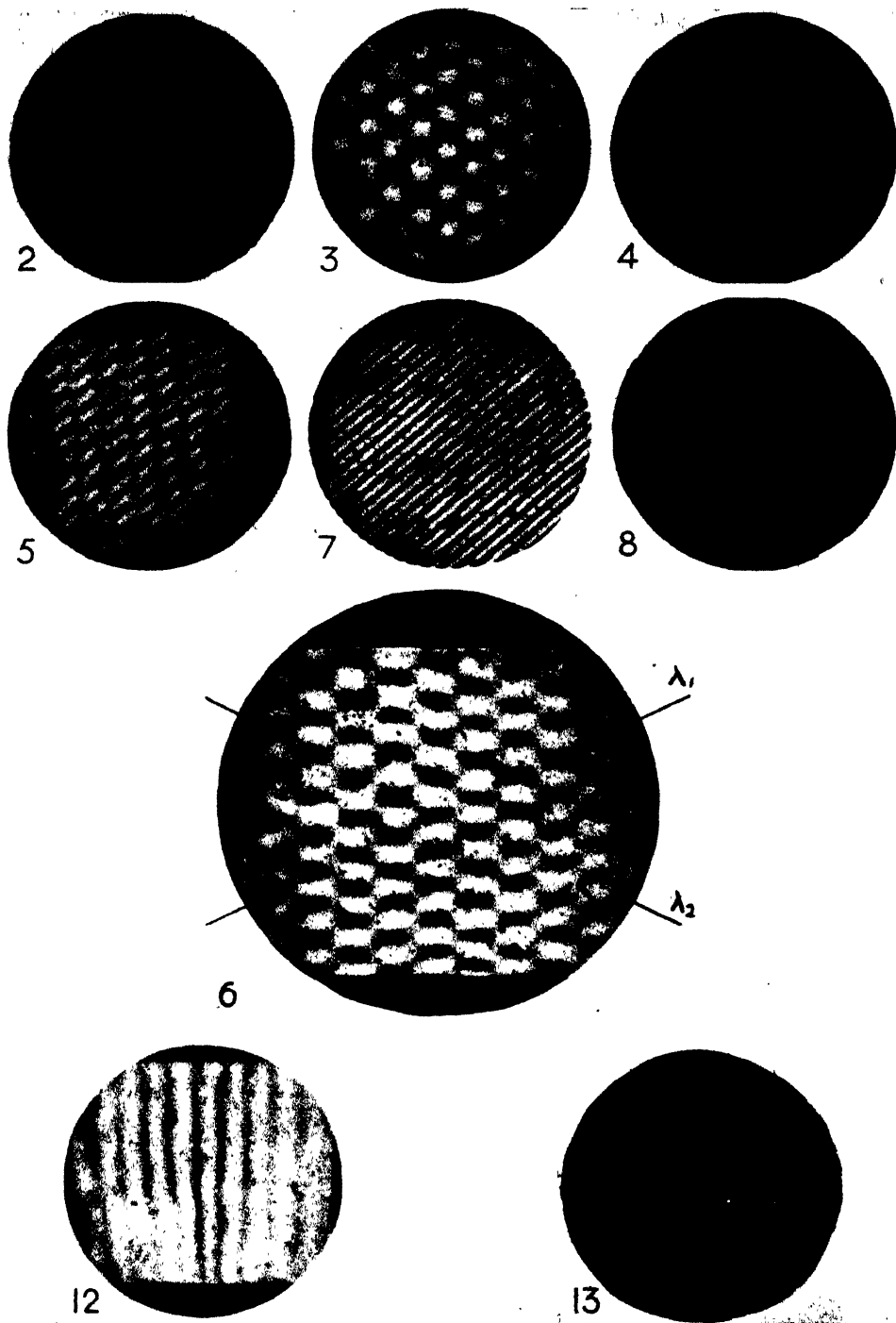


Figure 1.

prism of 30° angle was used, the arrangement being indicated in figure 1. For the light source *S* a hot cathode sodium lamp was employed in general, but occasionally it was replaced by a mercury vapour lamp with an absorption cell filled with potassium bichromate solution. The sodium yellow doublet has a separation of about 6 Å. and the mercury doublet a somewhat greater separation of 21·1 Å.



Figures 2-8 and 12 and 13.

Prism P is arranged at minimum deviation with respect to the rays emerging from the half-silvered plate T; the paths of the rays are indicated by the dotted lines. The initial adjustment is effected by viewing with an eyepiece the various reflected images of A formed in the focal plane E of lens L_2 : a single bright image is seen due to reflection from mirror M_1 , and a series of coloured images due to dispersion by prism P and reflection from mirror M_2 . By means of the mirror-adjusting screws the image from M_1 can be made to coincide with any one of these coloured images; with sodium light the two yellow images, from M_1 and M_2 respectively, are superimposed. Final adjustment is made in the usual way by placing the eye directly at E and moving M_1 longitudinally until the two paths are substantially equal, and then, by relative inclination of the mirrors, to produce a system of straight fringes.

Owing to the dispersion introduced by prism P two such sets of fringes are formed corresponding to the two radiations. These sets are parallel for one condition of adjustment only—when relative inclination of the mirrors is about an axis parallel to the refracting edge of the prism; a photographic record of the fringe system then formed is shown in figure 2. By suitable inclination of one of the mirrors the fringe systems can be arranged in any orientation; thus in figure 3 the systems are shown inclined to a horizontal line at angles of 39° and 143° respectively, in figure 4 at 12° and 170° , in figure 5 at 19° and 46° and in figure 6 at 23° and 154° . For comparison, in figure 7 is shown a system of parallel fringes obtained with the single radiation 5461 Å.

It is seen that the intersecting points of the two fringe systems lie on straight lines, whatever the orientation, and that a series of grey bands appear which are always vertical (that is, parallel to the refracting edge of the prism). These bands are the loci of those points where a maximum of one fringe system coincides with a minimum of the other system; they are shown most prominently in figure 2.

If the prism is of poor quality, due either to imperfection of its surfaces or to inhomogeneity of the glass, both fringe systems are curved considerably; the grey bands remain perfectly straight and parallel, however, as shown in figure 8. A piece of window glass interposed between the prism and mirror produces the same effect as an inferior prism.

The characteristics of these grey bands are: (i) the bands are independent of both the inclination and the shape of the interference fringes, always remaining straight and sensibly parallel to the refracting edge of the prism; (ii) for the same prism and the same radiations the spacing between the bands is constant; (iii) the spacing between the bands decreases as the difference in wavelength of the two radiations increases; (iv) a parallel displacement of one of the interferometer mirrors produces a movement of the bands. A linear relationship exists between mirror displacement and wavelength difference of the radiations: this was established and discussed in a previous paper (Tsien 1945).

§ 3. EQUATIONS FOR THE GREY BANDS

Single radiation

For a single radiation of wavelength λ and a small angle of inclination i between the interferometer mirrors the spacing a between two adjacent fringes is given by

$$a = \lambda/2i. \quad \dots\dots(1)$$

Orientation of the two fringe systems

Assume that the refracting edge of the prism is parallel to the z axis (figure 9), and that after refraction the normals to the wavefronts corresponding to the two neighbouring radiations, and designated by λ_1 and λ_2 , lie in the xy plane with λ_1 directed along the x axis and λ_2 inclined to it at an angle $\Delta\theta$.

Let the direction cosines of the normal to mirror M_2 be $\cos \alpha$, $\cos \beta$ and $\cos \gamma$, where α is a small quantity and both β and γ are very nearly equal to $\pi/2$. Mirror M_1 is assumed perpendicular to the single incident beam incorporating λ_1 and λ_2 .

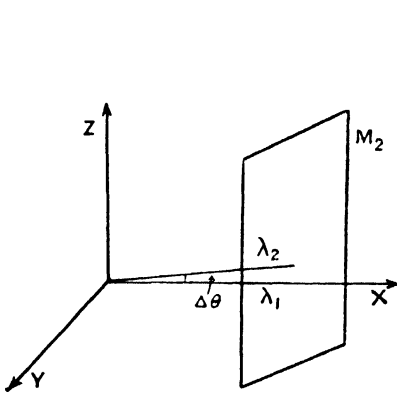


Figure 9.

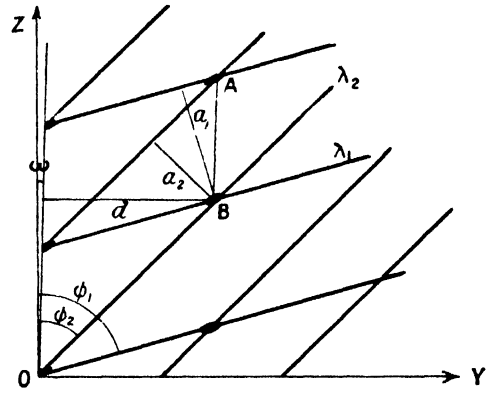


Figure 10.

First, consider the case of λ_1 . If the virtual intersecting angle between M_1 and M_2 is δ_1 , it is evident that $\delta_1 = \alpha$. The direction cosines of the intersecting line of the two mirrors will be

$$\cos \xi_1 = 0, \quad \cos \eta_1 = -\cos \gamma / \sin \delta_1, \quad \cos \zeta_1 = \cos \beta / \sin \delta_1.$$

The interference fringes of λ_1 will run in this direction.

Again, consider the case of λ_2 . The virtual intersecting angle δ_2 between the two mirrors will have the relation

$$\cos \delta_2 = \cos \alpha \cos \Delta\theta + \cos \beta \sin \Delta\theta. \quad \dots (2)$$

The direction cosines of the intersecting line of the two mirrors will be

$$\begin{aligned} \cos \xi_2 &= \sin \Delta\theta \cos \gamma / \sin \delta_2, & \cos \eta_2 &= -\cos \Delta\theta \cos \gamma / \sin \delta_2, \\ \cos \zeta_2 &= (\cos \Delta\theta \cos \beta - \sin \Delta\theta \cos \alpha) / \sin \delta_2. \end{aligned}$$

The interference fringes of λ_2 will run in this direction.

Projection of the fringes of λ_2 into the yz plane

Since we observe the pattern of the interference fringes in the yz plane, we must project the fringes on to that plane.

Let ϕ_1 and ϕ_2 be the angles made with the z axis by the lines of the two systems of fringes of λ_1 and λ_2 when projected on the yz plane.

It is evident that for λ_1 we have $\phi_1 = \zeta_1$, so that $\sin \phi_1 = -\cos \gamma / \sin \delta_1$, $\cos \phi_1 = \cos \beta / \sin \delta_1$, and for λ_2 we have $\cos \eta_2 : \cos \zeta_2 = \sin \phi_2 : \cos \phi_2$, so that $\sin \phi_2 = -(\cos \Delta\theta \cos \gamma) / \epsilon$, $\cos \phi_2 = (\cos \Delta\theta \cos \beta - \sin \Delta\theta \cos \alpha) / \epsilon$, where

$$\begin{aligned} \epsilon^2 &= \sin^2 \delta_2 - \sin^2 \Delta\theta \cos^2 \gamma \\ &= \cos^2 \Delta\theta \sin^2 \alpha + \sin^2 \Delta\theta \cos^2 \alpha - 2 \sin \Delta\theta \cos \Delta\theta \cos \alpha \cos \beta. \end{aligned}$$

Since α , $\Delta\theta$, δ_1 and δ_2 are small quantities, the above relation becomes

$$\epsilon^2 = \delta_2^2 = \alpha^2 + \Delta\theta^2 - 2\Delta\theta \cos \beta, \quad \dots\dots(3)$$

to the first order of approximation.

Here we may note that $\phi_2 = \zeta_2$ to the first order of approximation.

It is also useful to note the two following relations (to the first order of approximation),

$$\sin(\phi_2 - \phi_1) = (\Delta\theta \cos \gamma) / \delta_1 \delta_2, \quad \cos(\phi_2 - \phi_1) = (\delta_1^2 - \Delta\theta \cos \beta) / \delta_1 \delta_2.$$

Intersection of the two systems of fringes

Suppose that the origin is one of the intersection points of the two systems of fringes. Since we consider all fringes in the yz plane, we obtain the equation for the first system of fringes of λ_1 as

$$(y - ma_1 \cos \phi_1) / \sin \phi_1 = (z + ma_1 \sin \phi_1) / \cos \phi_1,$$

and that for the second system of fringes of λ_2 as

$$(y - na_2 \cos \phi_2) / \sin \phi_2 = (z + na_2 \sin \phi_2) / \cos \phi_2,$$

where m and n are positive and negative integers, including zero, and a_1 and a_2 are the respective fringe separations of the two systems.

For the point of intersection we find

$$y = (na_2 \sin \phi_1 - ma_1 \sin \phi_2) / \sin(\phi_1 - \phi_2), \quad z = (na_2 \cos \phi_1 - ma_1 \cos \phi_2) / \sin(\phi_1 - \phi_2).$$

By putting $m = n + k$ and letting n vary while keeping k as a constant integer, we find the equation for the system of lines formed by the intersection of these infinite number of points by eliminating n from the above two expressions, obtaining

$$\frac{y + ka_1 \sin \phi_2 / \sin(\phi_1 - \phi_2)}{a_2 \sin \phi_1 - a_1 \sin \phi_2} = \frac{z + ka_1 \cos \phi_2 / \sin(\phi_1 - \phi_2)}{a_2 \cos \phi_1 - a_1 \cos \phi_2}. \quad \dots\dots(4)$$

This represents a system of parallel lines as k varies integrally, corresponding to the loci of the points of intersection of fringes of the two systems, maxima coinciding to give rows of points of high illumination (seen as finite areas), minima coinciding to give rows of dark points, and maxima of one system coinciding with minima of the other to give the grey bands.

Verticality of the bands

Suppose that the bands are inclined at an angle ω to the z axis which is the direction of the refracting edge of the prism P, then

$$\sin \omega = \frac{a_2 \sin \phi_1 - a_1 \sin \phi_2}{\sqrt{\{a_1^2 + a_2^2 - 2a_1a_2 \cos(\phi_1 - \phi_2)\}}}. \quad \dots\dots(5)$$

Since by (1) $a_1 = \lambda_1 / 2\delta_1$ and $a_2 = \lambda_2 / 2\delta_2$, we obtain, using the relation of equation (3),

$$\sin \omega \simeq \frac{(\lambda_2 - \lambda_1) \cos \gamma}{\sqrt{\{\lambda_1 \lambda_2 (\Delta\theta)^2 + (\lambda_2 - \lambda_1)(\lambda_2 \delta_1^2 - \lambda_1 \delta_2^2)\}}} \simeq \frac{(\lambda_2 - \lambda_1) \cos \gamma}{\Delta\theta \sqrt{(\lambda_1 \lambda_2)}} \simeq \frac{(\lambda_2 - \lambda_1) \cos \gamma}{\lambda \Delta\theta}. \quad \dots\dots(6)$$

Although $\cos \gamma$ can be varied at will, γ is always very near to $\pi/2$, so that ω varies very slightly and is practically equal to zero, for $\lambda_2 - \lambda_1 = \Delta\lambda$ is negligible compared with λ . When the two systems of fringes are vertical and parallel, $\cos \gamma = 0$ and $\omega = 0$ exactly. This is the only condition for the exact verticality of the bands. On the other hand, it is not difficult to see that the two systems of fringes cannot be horizontal at the same time.

For a rough estimate of the magnitude of ω , take the case of the mercury yellow doublet with a prism of light flint, and assume that γ deviates about $1'$ from 90° . It is found that ω is about $10'$.

Spacing of the bands

From geometry the relation for the spacing of the bands is

$$d = a_1 a_2 / \sqrt{a_1^2 + a_2^2 - 2a_1 a_2 \cos(\phi_1 - \phi_2)}.$$

By using this formula to the first order of approximation, and making the same substitutions as were used in deriving equation (6) from equation (5), we find

$$d \simeq \sqrt{(\lambda_1 \lambda_2) / 2\Delta\theta} \simeq \lambda / 2\Delta\theta. \quad \dots\dots(7)$$

It will be noted that the mathematical form of equation (7) is exactly the same as that of equation (1). Comparing the physical meaning of these two equations, we see that the spacing of the bands is equal to the spacing of one system of fringes when the other system has been made null. The above statement has been proved experimentally. After a few trials, no difficulty is found in adjusting mirror M_2 so that it is perpendicular to one radiation. Figure 12* is a photograph taken under these conditions. The fringes are not very straight owing to the imperfect quality of the prism. It is also obvious that under this condition the interference fringes must be parallel to the refracting edge of the prism, if the quality of the latter is fairly good. For when M_2 is perpendicular to λ_1 , we have $\cos \alpha = 1$, $\cos \beta = 0$, $\cos \gamma = 0$, and therefore $\sin \phi_2 = 0$.

§ 4. DISPERSION

It is apparent from equation (7) that the dispersion ($d\theta/d\lambda$) of the prism in use can be determined by measuring the spacing of the grey bands. Such a determination was made for various prisms, and the results obtained were in good agreement with those based on a calculation using Hartman's dispersion formula.

§ 5. THE RESOLUTION OF $H\alpha$

The cross-pattern of interference fringes may offer a new contribution to the resolution of spectral lines. The existence of bands formed by the intersection of the two systems of fringes serves as a good indication of the success in resolution. The spacing between two bands gives the dispersion of the prism for these radiations, and the shifting of the bands as one of the reflecting mirrors is displaced yields the value of the difference of the two radiations in question (Tsien 1945). For convenience in measurements, the spacing d between two bands is brought to a suitable width, from 1 mm. to a few centimetres. For a large difference between the wavelengths of two radiations, a prism of low dispersion glass and small refracting angle is chosen, and conversely, for a small difference between the wavelengths a prism of high dispersion glass and large refracting angle is preferable.

* Figure 11 omitted during amendment.

An attempt to resolve the red Balmer hydrogen line ($H\alpha = 6562.82 \text{ \AA}$.) was made. It has been known as a doublet with a separation of about $0.13 \text{ \AA}$. The source of light was a water-cooled hydrogen discharge tube of low pressure excited at 2000 v. A 65° prism made of extra dense flint and of 8 cm. side was used; owing to the large incident angle at the prism surface, however, the aperture of the prism is considerably reduced. A rough estimate of the dispersion $d\theta/d\lambda$ of the prism at this radiation gives about 3000 radians per centimetre; then, by calculation, the spacing of the bands in this arrangement is about 8.4 cm. Thus in the experiment two bands cannot be seen simultaneously. In spite of this, the shifting of a single band should be detectable, as one of the reflecting mirrors makes a parallel displacement. According to the equation suggested in the previous paper (Tsien 1945), the displacement required to shift a band by one spacing is about 1.6 cm. As a matter of fact, in this experiment the interference fringes disappeared as the mirror made a displacement of only 4 or 5 mm. In a paper by Buisson and Fabry (1912) it was pointed out that the two components of $H\alpha$ from a hydrogen discharge tube at room temperature are large enough to overlap each other; if the tube were put into liquid air, they would be perfectly separated. This is due to the Doppler effect which predominates in a light gas at ordinary temperature. As no liquid air was available, this method could not be explored.

In order to reduce the spacings of the bands and make the intersection of the two systems of fringes visible, it is necessary to increase the dispersion of the prism. There are three ways of doing this. Firstly, increasing the number of prisms employed is certainly beneficial; it is preferable to put a second prism in the other beam of light so as to make the absorption in the two beams equal. However, the use of prisms in succession reduces the effective aperture of the prism. Secondly, the dispersion is increased by increasing the refracting angle of the prism; the consequent reduction of aperture is, however, accompanied by a loss of light. A full discussion of this point is given in E. C. C. Baly's *Spectroscopy*, pp. 66-72. Generally speaking, a prism of refractive index below 1.7 and refracting angle not more than 65° is suitable. Thirdly, a glass of high refractive index usually possesses high dispersive power, $dn/d\lambda$. Wood (1934) has discovered the remarkable optical properties of nitro-dimethyl-aniline. It has an enormous dispersive power in the yellow and green regions, the term $dn/d\lambda$ near $H\alpha$ being about 6600 cm^{-1} . If a 60° prism of such material is used, the separation of bands for the $H\alpha$ doublet can be reduced to no more than 2 cm. Moreover, imperfections in the optical quality of the organic prism, whilst being a serious drawback to its use in a spectrograph, in no way affect the formation of the grey bands.

§ 6. DISTORTION OF THE FRINGES

Since λ_1 and λ_2 differ by a very small amount only, any local inhomogeneity in the prism introduces substantially the same change in optical path, and this produces the same fractional fringe shift. Thus at a point of intersection of any two fringes of the two systems, in a region affected by the prism defect, each fringe is displaced the same fractional part of the corresponding fringe separation. By simple geometry, the point of intersection must be displaced along the line of intersection points on which the undisplaced point lies. Therefore the lines of intersection, and thus the grey bands, are unaffected by prism defects as regards both homogeneity and surface flatness.

ACKNOWLEDGMENTS

It is the author's pleasure to express his thanks to Professor E. N. da C. Andrade, F.R.S., for his interest in the paper and to Mr. K. J. Habell for his work in revising the draft and for many valuable suggestions. His thanks are also due to Mr. D. Orson Wood, who has read through the paper, and to Dr. Ny Tsi-Zé and Dr. J. S. Wang for their useful discussions.

REFERENCES

- BUISSON, H., and FABRY, CH., 1912, *J. Phys. Radium*, **2**, 442.
 TSIEN, L. C., 1945, *Chinese J. Phys.*, **6**, 67.
 TWYMAN, F., 1918, *Phil. Mag.*, **35**, 49.
 WOOD, R. W., 1934, *Physical Optics*, 3rd edition (New York: Macmillan Co.), p. 508.

A Method of Fitting the Andrade Creep Equation to Experimental Results

By A. J. KENNEDY

University College, London

Communicated by E. N. da C. Andrade; MS. received 24 May 1948

ABSTRACT. A method is described whereby the constants in Andrade's formula, $l = l_0(1 + \beta t^{\frac{1}{2}})e^{\kappa t}$, for the flow of metals under constant stress, can be rapidly deduced from experimental results by direct reading from a system employing sliding templates of calculated shape. The general equation to which the method is applicable is deduced.

§ 1. INTRODUCTION

IT is now generally accepted (for example Cottrell and Aytakin 1947, Orowan 1947, Latin 1948) that Andrade's formula $l = l_0(1 + \beta t^{\frac{1}{2}})e^{\kappa t}$ represents very closely the creep of metals under constant stress, and effects a separation of the β , or transient, flow, and the κ , or permanent, flow, where each has a distinct physical significance (Andrade 1910, 1914, Andrade and Chalmers 1932). The formula, however, is a somewhat troublesome one to fit to experimental values, and in the past the method employed has been either that of determining β for small values of t , where the κ -term is small, and hence calculating κ from the results for a larger value of t , with suitable subsequent adjustment, or one of the "algebraic" methods which depend on the selection of particular times in order that an equation involving only one of the constants may be formed (the use of three times separated by equal intervals, for instance, provides a quadratic in β). All the methods are lengthy, and particularly tedious if a large number of curves have to be fitted. In the following an account is given of a method of obtaining values of β , κ and l_0 , from any creep curve, by direct graphical reading.

§ 2. THEORY OF THE METHOD

1. It is convenient to put $l_0 = f_1$, $(1 + \beta t^{\frac{1}{2}}) = f_2$ and $e^{\kappa t} = f_3$, and to use the notation $()_1, ()_2, \dots$ to represent values at $t_1, t_2, \dots$. Consider the curve $(f_2 f_3)_1 = l/l_0 = L$, where κ and β are treated as the variables. We require to know what effect altering the value of L has on the shape of the curve.

Differentiating we obtain $-Lt_1 e^{-\kappa t_1} (d\kappa/d\beta) = t_1^{\frac{1}{2}}$ or $d\kappa/d\beta = -t_1^{-\frac{1}{2}}/(1 + \beta t_1^{\frac{1}{2}})$, and $Lt(d\kappa/d\beta) = -t_1^{-\frac{1}{2}}$.

As the slope of this κ - β curve for a constant t depends only on β , then altering the value of L merely shifts the curve in a direction parallel to the κ -axis, its shape remaining the same. It remains to find how the shift is related to changes in the value of L . For $\beta=0$, and for the two curves $(f_2f_3)_1=M$ and $(f_2f_3)_1=N$, we have $\kappa_M=t_1^{-1}\log M$ and $\kappa_N=t_1^{-1}\log N$. Hence we may obtain the curve $(f_2f_3)_1=N$ by displacing $(f_2f_3)_1=M$ a distance corresponding to $t_1^{-1}\log(M-N)$, or, in other words, if a vertical logarithmic scale is added for the $M, N \dots$ values, then equal shifts with respect to this scale correspond to equal changes in the value of the L term.

2. The curves $(f_2f_3)_1=M$ and $(f_2f_3)_2=N$, will provide, at their point of intersection, values of β and κ which will satisfy both curves, that is, fulfil the required conditions at both t_1 and t_2 . It is clear that the curves $(f_1f_2f_3)_1=M'$ and $(f_1f_2f_3)_2=N'$ are also of the same general form, the effect of change in the f_1 term being to displace the curve in a direction parallel to the κ -axis, just as change in the value of the M' term does. We require next to know how this point of intersection varies with change in the f_1 value. This may be obtained from

$$\frac{M'}{N'} = \frac{(1 + \beta t_1^{\frac{1}{2}})}{(1 + \beta t_2^{\frac{1}{2}})} \cdot e^{(t_1 - t_2)\kappa}$$

for, differentiating,

$$\frac{M'}{N'} (t_2 - t_1) e^{(t_1 - t_2)\kappa} \frac{d\kappa}{d\beta} = (t_1^{\frac{1}{2}} - t_2^{\frac{1}{2}}) (1 + \beta t_2^{\frac{1}{2}})^{-2}$$

which reduces to $(1 + \beta t_1^{\frac{1}{2}})(1 + \beta t_2^{\frac{1}{2}})(t_2 - t_1)(d\kappa/d\beta) = (t_1^{\frac{1}{2}} - t_2^{\frac{1}{2}})$. When $\beta=0$, $d\kappa/d\beta = (t_1^{\frac{1}{2}} - t_2^{\frac{1}{2}})/(t_2 - t_1) = C$, a constant, for all values of f_1 , once the times t_1 and t_2 have been chosen. At any other point, $d\kappa/d\beta = C$ multiplied by a function of β only. Therefore, as f_1 varies, the locus of the point of intersection of the curves $(f_1f_2f_3)_1=M'$ and $(f_1f_2f_3)_2=N'$ is of a determined shape, independent of the values of M' and N' , but displaced parallel to the κ -axis an amount determined by the choice of their values (see figure 1). If a similar locus is obtained for times t_2 and t_3 (the time t_3 corresponding to a length O'), it is apparent that the point of intersection of the two loci represents a point of common values for β , κ and l_0 for the three curves

$$(f_1f_2f_3)_1=M', (f_1f_2f_3)_2=N'$$

and $(f_1f_2f_3)_3=O'$.

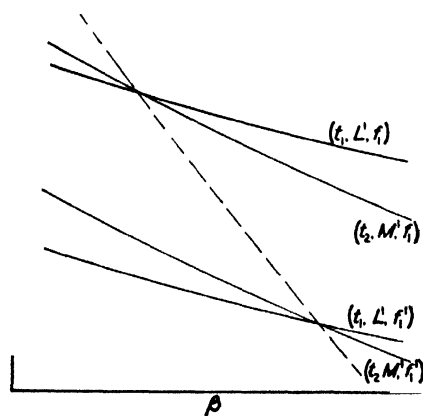


Figure 1.

--- Locus of point of intersection of t_1 and t_2 curves as f_1 varies.

§ 3. METHOD

Using the foregoing theory, we proceed to a suitable graphical method for finding the values of the constants β , κ and l_0 from experimental curves. Graphs are plotted, on a suitably large scale, of κ against β for the three particular times selected: it has been shown that the *shape* of any one of these is fixed once t is fixed. From these curves, templates are cut from some suitable material, such as Bristol board. The method, then, depends upon finding some point at which

all three curves will intersect (hence providing the same values for κ and β) and at which each template will indicate the same l_0 . The next step in the construction is to mark each template with a logarithmic scale (AA' in figure 2). This is best carried out by setting the extreme left of the template, i.e. the point X (figure 2) against those calculated values of κ which give $(f_2 f_3) = 1.01, 1.02, 1.03, \dots 1.1 \dots 1.2 \dots$ etc., and inscribing these values on the template in relation to some fiducial mark Q. The three fiducial marks (one for each template) can, of course, be placed anywhere, but if these are as widely separated as possible, then the subsequent laying-out of the l_0 -scale (described later) is made easier. If we now assign to these three fiducial marks a suitable round number for l_0 , corresponding approximately to the original length of the experimental specimens (the value 20 cm. will be used throughout in the subsequent explanation), then each mark may be spread into logarithmic scales the points on

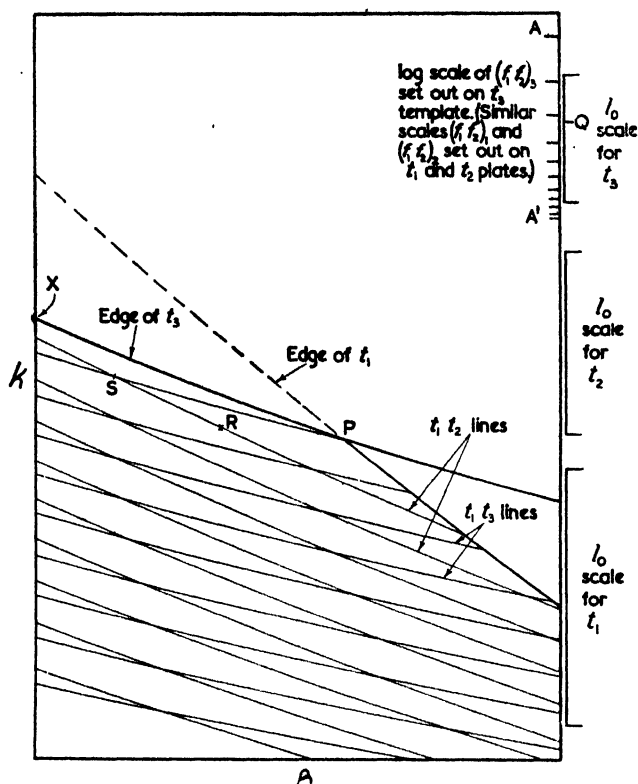


Figure 2.

Diagram shows the t_1 and t_3 templates in position providing the point P. If the setting of the t_1 and t_2 templates provides the point R then the point S is identifiable as that point for which values l_0 , κ and β for the times t_1 , t_2 and t_3 will coincide.

which correspond to various values of l_0 . These scales are easily laid out, the distance between say, 20.0 and 21.0 cm. being the distance between 1.00 and 1.05 on the respective template scale, and are preferably marked on a raised strip fixed to the card or board on which the diagram is drawn, so that the templates may be conveniently slid against this in a direction parallel to the κ -axis.

Values of l at times t_1 , t_2 and t_3 are read off from the experimental curve and divided by 20 to give $(f_2 f_3)_1$, $(f_2 f_3)_2$ and $(f_2 f_3)_3$. Any l_0 value is chosen (20 cm. for instance) and the templates are then so placed that the respective $(f_2 f_3)$ values are each set against this value of l_0 . The curves will, in general, intersect in a small triangle, which may, by variation of the value of l_0 chosen, be reduced to a point, or at least to a minimum area. This would provide the answer we require, but by the use of the second part of the theory, if a permanent curvilinear grid is laid out on the graph, one set of lines corresponding to convenient displaced positions of the $t_1 t_2$ locus and the other set to similar positions of the $t_2 t_3$ locus, then the point required is identifiable by eye as the point of intersection of the

two grid lines that pass through the points given by the intersection of the t_1 and t_2 , and the t_2 and t_3 templates.

The "grid" may be obtained by taking the t_1 and t_2 templates and marking on the graph their point of intersection when each is set, with reference to its corresponding l_0 scale, against the same value of l_0 . Points are plotted for various values of l_0 , and the curve obtained gives the shape of all such loci for these two times. Similar curves are drawn by displacing this curve through suitable regular intervals in a direction parallel to the κ -axis. The procedure is repeated for two other templates, say t_2 and t_3 , and the required grid is thus obtained. Figure 2 shows how the point required, S, is identified, and this point may now be "tested" for coincidence of all three templates; the experimental ($f_2 f_3$) values should now indicate the same l_0 on their respective scales, and the value of l_0 may therefore be read off from any one of them.

The grid lines themselves, of course, require no scaling, and by a judicious choice of the times t_1 , t_2 and t_3 their slopes can be arranged to give an open and workable lattice.

There are many ways in which the use of a "grid" to locate the point required may be applied. It may, for instance, be laid out on a separate sheet, the two points provided by the initial settings of the plates being plotted on this grid and the common point identified as before. The most satisfactory method so far devised is that which uses a transparent sheet, inscribed, as in figure 3, with two lines from one "family" and one from the other (it will be evident that the point for the times t_1 and t_2 might be above or below, to the left or to the right of that for the times t_2 and t_3), the plate being moved laterally and vertically, that is, without rotation, into such a position that the $t_1 t_2$ and $t_2 t_3$ points lie on their respective loci. This is shown in figure 3.

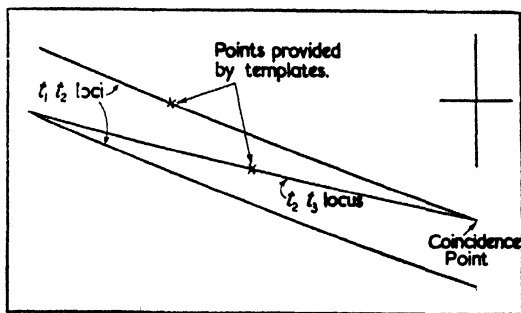


Figure 3.

Whichever method is used, the point obtained is checked by the templates and the value of l_0 read off. The value obtained may, of course, be either greater or less than that of the original setting.

§ 4. THE GENERAL EQUATION TO WHICH THE METHOD IS APPLICABLE

Consider the line given by the intersection of the surfaces

$$f(x, y, z, a) = A ; \quad f(x, y, z, b) = B. \quad \dots\dots (1)$$

(The variables x , y and z in these generalized equations are equivalent to l_0 , β and κ in the equation considered in the previous sections, a and b being equivalent to values of time, and A and B equivalent to values of length obtained experimentally.)

If the method discussed is to be applicable to this function, then firstly the projection of the line of intersection of two such functions on to the xy plane must be a curve whose shape remains unchanged by change in the value of A or B . This curve in the xy plane is equivalent to the locus of the point of intersection of two templates as l_0 varies. Secondly, the line of intersection of $f(x, y, z, a) = A$

with a plane $z=k$, a constant, must also be of such a form that its shape is unchanged by change in the value of A . This curve corresponds to the outline of the template itself.

Eliminating z from equation (1) we obtain the cylinder generated by lines parallel to the z -axis which pass through the line of intersection of the surfaces. The projection of this line on to the xy plane is given by the intersection of the cylinder and $z=0$, that is,

$$\phi(x, y, a, b) = \chi \text{ and } z=0. \quad \dots\dots(2)$$

If this curve is to have the same shape for all values of χ , a change of the χ -value merely translating it along the y -axis, then it must be of the general form $y = F(x) + \chi$ where χ is an independent variable. Or, more generally, the required family may be written

$$\omega\{y - F(x)\} = \text{constant} \quad \dots\dots(3)$$

where the constant has any assigned value. (The function $\omega(t) = \text{constant}$ also has the solution $\omega(t) = t^n$ as $n \rightarrow 0$. This, whilst it fulfils the condition, is not one to which the method may be applied, as the change in the values of the parameters with change in the values of the applied conditions becomes too small to be investigated by such a means.)

We may now deal in a similar manner with the second condition, that putting $z=k$, a constant, in the equation $f(x, y, z, a) = A$ provides a curve with similar properties for change in the value of A .

This is obviously fulfilled (as before) by an equation of the form

$$\delta\{y - \mathcal{F}(x)\} = \text{constant}. \quad \dots\dots(4)$$

If a curve of this form is to result from putting $z = \text{constant}$ in either of equations (1), then the f -function must be of such a form that it may be written as $\delta\{y - \mathcal{F}(x)\} = \eta(z)$, for only if the z -function is separable does the function of $y - \mathcal{F}(x)$ remain unchanged for change in z , providing, for all functions and values of z , a family of curves with the required property.

Applying this, we may therefore write (1) in full as

$$\lambda\{y - F(x, a)\}\rho(z, a) = A; \quad \lambda\{y - F(x, b)\}\rho(z, b) = B \quad \dots\dots(5)$$

(retaining a, b, A and B in the notation).

If, now, our other condition is to apply, that is, that embodied in (3), the elimination of z between these two equations must provide a curve which may be written $\omega\{y - F(x)\} = \text{constant}$.

But the elimination of z from (5) provides, in general,

$$\Omega\{y - F(x, a), y - F(x, b), a, b\} = \text{constant}$$

(assuming we can obtain z on one side of at least one of the equations). If this is to be expressible in the form of (3), then either $F(x, a) = F(x, b)$ giving $a = b$, which is experimentally impossible, or, the elimination of z provides an equation of the form

$$\xi\{y - f_1(x)\} = \text{constant} \times \xi\{y - f_2(x)\} \quad \dots\dots(6)$$

which can always be reduced to the form of (3).

The function of z , $\rho(z, a)$, must therefore be separable as $\mu(z)\psi(a)$, so that for all functions and values of these the curve given by (5) is reducible to that of (6) and hence to that of (3).

The general equation of the surface to which the method applies may therefore be fully stated as

$$\mathcal{F}_1\{y - f(x, a)\}\mathcal{F}_2(z)\mathcal{F}_3(a) = A. \quad \dots\dots(7)$$

It follows, therefore, that if there are two quantities (a and A) between which an experimental relation has been formed, and that this relation can be expressed by an equation involving three constants x , y and z , then the method described can be used to find the values of the constants from any three readings on condition that (i) one of the constants is separable from the others and also from the experimental quantities, and (ii) that the remaining constants may be expressed in the form $\omega\{y - f(x)\}$.

ACKNOWLEDGMENT

I am sincerely indebted to Professor E. N. da C. Andrade, F.R.S., for his considerable advice and interest, and also to Dr. R. Eisenschitz for discussion with him on the generality of the method.

REFERENCES

- ANDRADE, E. N. DA C., 1910, *Proc. Roy. Soc. A*, **84**, 1; 1914, *Ibid.*, **90**, 392.
 ANDRADE, E. N. DA C., and CHALMERS, B., 1932, *Proc. Roy. Soc. A*, **138**, 348.
 COTTRELL, A. H., and AYTEKIN, V., 1947, *Nature, Lond.*, **160**, 328.
 LATIN, A., 1948, *J. Inst. Metals*, **74**, 259.
 OROWAN, E., 1947, West of Scotland Iron and Steel Institute.

The Influence of a Transverse Magnetic Field on a Cylindrical Plasma

By LARS BECKMAN

Department of Electronics, The Royal Institute of Technology, Stockholm

Communicated by A. von Engel; MS. received 16 March 1948

ABSTRACT. The conditions in a cylindrical plasma under the influence of a transverse homogeneous magnetic field are treated by known methods. It is shown that the magnetic field deflects the column towards the wall with the result that the total loss of electrons and ions is increased. This causes an increase in electron temperature and in the axial field strength. Formulae relating the axial electric field, the magnetic field and the electron temperature are derived as well as equations describing the electron density distribution. Some measurements have been made in order to test the theory. The agreement is fair, taking account of the limits of experimental accuracy and of uncertainty in the constants used in the calculations.

§1. INTRODUCTION

THE conditions in a cylindrical plasma under the influence of a transverse homogeneous magnetic field are treated in a manner analogous to that given by von Engel and Steenbeck (1934, pp. 82-87) for the case without magnetic field. The expressions for the electron drift in a magnetic field used here have been given by Tonks and Allis (1937).

The case to be treated is that in which the plasma is limited to a cylindrical tube and the current flowing in the direction of the axis. When a homogeneous magnetic field is applied at right angles to the axis the current is forced towards the tube wall thereby increasing the loss of electrons and ions. In order to compensate this loss the axial electron field increases, thus increasing the ionization and the electron temperature.

§ 2. THEORY

In the steady state the conditions in the plasma can be expressed by the equation

$$\operatorname{div}(Nv) = z'N, \quad \dots\dots(1)$$

where N denotes the electronic or ionic density, v the drift velocity of the electrons and ions, and z' the average number of ion pairs created by one electron per unit of volume and time. Equation (1) will be valid under the assumptions that the electronic and ionic densities are approximately equal and that the recombination of ions and electrons and the formation of negative ions can be neglected.

A rectangular coordinate system is assumed with the z axis along the axis of the cylinder with its positive direction in the direction of the current. The magnetic field is in the y direction. Expression for the electron drift velocity components are given by Tonks and Allis (1937, equations (10) and (15)), viz.

$$\left. \begin{aligned} v_{x_e} &= -\alpha \left(b_e E_x + \frac{D_e}{N} \frac{\partial N}{\partial x} \right) - \beta b_e E_z, \\ v_{y_e} &= -b_e E_y - \frac{D_e}{N} \frac{\partial N}{\partial y}, \quad v_{z_e} = -\alpha b_e E_z + \beta b_e E_x. \end{aligned} \right\} \dots\dots(2)$$

Here E_x , E_y and E_z are the components of the electric field, b_e and D_e are coefficients of mobility and diffusion, N is the electron density and α and β are dimensionless quantities defined by Tonks and Allis:

$$\left. \begin{aligned} \alpha &= 1 - h^2 + h^4 \exp h^2 \int_h^\infty \frac{\exp -h}{h} dh, \\ \beta &= \frac{1}{2} h \sqrt{\pi} (1 - 2h^2 + 4h^3 \exp h^2 \int_h^\infty \exp -h^2 dh), \\ h &= eB\lambda/mw, \end{aligned} \right\} \dots\dots(2a)$$

where B is the magnetic field, λ the electron mean free path and w is the most probable electron speed which is given in terms of the electron temperature T by

$$w^2 = 2kT/m.$$

The electron velocity distribution is supposed to be Maxwellian.

The corresponding expressions for the ionic velocity components are

$$\begin{aligned} v_{x_i} &= b_i E_x - \frac{D_i}{N} \frac{\partial N}{\partial x} - b_i^2 B E_z, \\ v_{y_i} &= b_i E_y - \frac{D_i}{N} \frac{\partial N}{\partial y}, \quad v_{z_i} = b_i E_z, \end{aligned}$$

because in this case h is very small and so $\alpha = 1$ and $\beta = b_i B$. For v_{z_i} the term $\beta b_i E_x$ is much smaller than $b_i E_z$ and so will be neglected.

In the steady state $v_{x_e} = v_{x_i} = v_x$ and $v_{y_e} = v_{y_i} = v_y$; we can eliminate E_x and E_y and the following expressions are obtained:

$$\left. \begin{aligned} v_x &= -\frac{1}{N} \frac{\partial N}{\partial x} \frac{b_i D_e + b_e D_i}{b_i + \alpha b_e} - \alpha - b_e b_i E_z \frac{\beta + \alpha b_i B}{b_i + \alpha b_e}, \\ v_y &= -\frac{1}{N} \frac{\partial N}{\partial y} \frac{b_i D_e + b_e D_i}{b_i + b_e}. \end{aligned} \right\} \dots\dots(3)$$

Since $\alpha b_e \gg b_i$ and

$$\frac{b_i D_e + b_e D_i}{b_i + b_e} \simeq \frac{D_e}{b_e} b_i = \frac{kT}{e} b_i = D_a \quad \dots\dots(4)$$

the coefficient of ambipolar diffusion, we obtain

$$v_x = -\frac{D_a}{N} \frac{\partial N}{\partial x} - c, \quad v_y = -\frac{D_a}{N} \frac{\partial N}{\partial y},$$

where

$$c = b_i E_z b_e \frac{\beta + \alpha b_i B}{b_i + \alpha b_e} \simeq b_i E_z \frac{\beta}{\alpha} \quad \dots\dots(5)$$

Equation (1) can now be written

$$D_a \left(\frac{\partial^2 N}{\partial x^2} + \frac{\partial^2 N}{\partial y^2} \right) + c \frac{\partial N}{\partial x} + z' N = 0.$$

After substituting $N = N' \exp(-cx/2D_a)$ and passing to cylindrical coordinates r and ϕ , the solution is obtained

$$N = N_0 \exp(-cr \cos \phi / 2D_a) J_0 \left\{ r \sqrt{\left(\frac{z'}{D_a} - \frac{c^2}{4D_a^2} \right)} \right\}, \quad \dots\dots(6)$$

where N_0 is the density along the z axis. Let $N=0$ approximately at the wall, thus

$$r_0 \sqrt{\left(\frac{z'}{D_a} - \frac{c^2}{4D_a^2} \right)} = 2.405 \dots = n. \quad \dots\dots(7)$$

r_0 is the tube radius and n the first zero of the Bessel function J_0 . Equation (6) can be rewritten

$$N = N_0 \exp(-cr \cos \phi / 2D_a) J_0(nr/r_0).$$

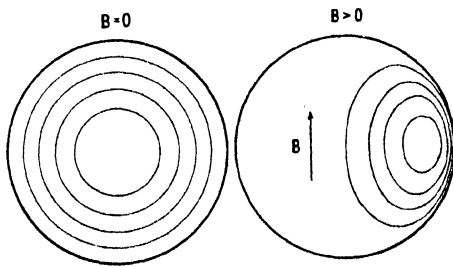


Figure 1.

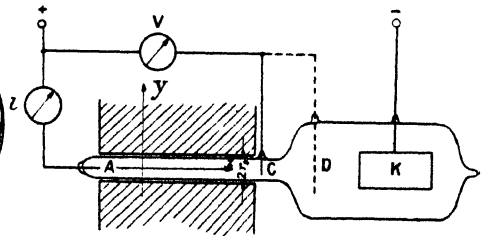


Figure 2.

This equation describes the electron density distribution. This is shown in figure 1 where the lines representing $N=\text{constant}$ are drawn in the cross section of the tube.

Now it is possible to express E_z and B as functions of the electron temperature T . According to von Engel and Steenbeck (1932, pp. 88–89) the rate of ionization of an electron is

$$z' = \frac{2amp}{e\sqrt{\pi}} \left(\frac{2kT}{m} \right)^{3/2} \exp(-eV_i/kT) \left(1 + \frac{1}{2} \frac{eV_i}{kT} \right), \quad \dots\dots(8)$$

where V_i is the ionization potential and p the pressure in mm. Hg. Equation (8) is the same as their equation (115) except for the factor 2 (instead of 600) as we are

using here m.k.s. instead of c.g.s. units. Equation (8) is derived under the simplifying assumption that the electron mean free path is independent of the electron temperature. Values of the constant a are also given by von Engel and Steenbeck.

From (2) and (3) we obtain the axial component of the electron velocity

$$v_{z_0} = -b_e E_z \left(\alpha + \frac{\beta^2}{\alpha} - \frac{\beta^2 b_1 / \alpha + \beta b_1^2 B}{b_1 + \alpha b_e} \right) + \frac{\beta}{N} \frac{\partial N}{\partial x} \left(\frac{D_a}{\alpha} - D_e \right).$$

Here $\beta^2 b_1 / \alpha + \beta b_1^2 B = (1 + \alpha b_1 B / \beta) \beta^2 b_1 / \alpha$ and it is obvious that $\alpha b_1 B / \beta \ll 1$ since for very strong magnetic fields α / β approaches zero. The last term in v_{z_0} is approximately equal to

$$-\beta \frac{D_e}{N} \frac{\partial N}{\partial x} = -\frac{\beta}{N} b_e \frac{kT}{e} \frac{\partial N}{\partial x}$$

and it is seen that this term can be neglected if $E_z \gg (\beta kT / Ne) \partial N / \partial x$. Thus by assuming as before $\alpha b_e \gg b_1$, we obtain $v_{z_0} = -b_e E_z (\alpha + \beta^2 / \alpha)$, which means that in the presence of a transverse magnetic field the drift velocity is decreased by a factor $(\alpha + \beta^2 / \alpha)$, the numerical value of which is 1.0 for zero magnetic field and then decreases when the magnetic field increases towards the value 0.883.

The electric field as a function of the electron temperature at zero field is, according to von Engel and Steenbeck (1932, pp. 183–185), but with regard to the above expression for the axial velocity,

$$E_z \left(\alpha + \frac{\beta^2}{\alpha} \right)^{1/2} = \frac{2\sqrt{2}}{\pi^{1/2}} \frac{\sqrt{\kappa}}{p\lambda} \frac{kT}{e} p. \quad \dots\dots(9)$$

Here κ is the electron's average relative loss of energy by an impact with a molecule. The electron temperature T is very strongly affected by the magnetic field.

The equations (4), (5), (7)–(9) form a system from which the relation between E_z and B can be calculated. For the sake of simplicity we introduce a new variable u defined by the equation $u = \sqrt{(eV_1 / kT)}$. Then we obtain

$$\frac{\beta^2}{\alpha(\alpha^2 + \beta^2)^2} = f(h) = \frac{1}{\kappa} \left[2 \sqrt{\left(\frac{2eV_1}{m} \right) \frac{a(p\lambda)^2}{(pb_1)}} u \exp(-u^2) (u^{-2} + \frac{1}{2}) - \frac{2.88\pi(p\lambda)^2}{(pr_0)^2} \right]. \quad \dots\dots(10)$$

From (2a) follows that $f(h)$ is approximately proportional to h^2 . Finally we express h in terms of u and B

$$h = -\lambda u B \sqrt{(e / 2mV_1)}. \quad \dots\dots(11)$$

The right-hand side of (10) is a function of u since κ is a function of u . For a certain value of u we can calculate $f(h)$ and then h . Then equation (11) gives the corresponding value of B and equation (9) that of E_z .

§ 3. EXPERIMENTS

Some measurements have been made on cold cathode glow discharges in different gases. Figure 2 shows the experimental arrangements.

The tubes were provided with a hollow cathode K, an anode A and a third electrode placed in one of the two positions C or D. According to this position the tubes will be divided into two groups referred to as type C and D. The tube was placed between the pole-pieces of an electromagnet. The distance between the anode and the third electrode is d and the length of the pole-pieces l .

Gas	p (mm. Hg)	Tube type	r_0 (cm.)	d (cm.)	l (cm.)	Anode current (ma.)	$E_{z0} = \frac{V_0}{d}$ (v/cm.)	Calculated values of E_{z0} (v/cm.)	T_0 °K.	Fig. No.
H ₂	5.0	C	0.3	3.5	3.5	15	71.5	74.5	18400	3
H ₂	2.7	D	1.0	6.0	4.5	10	33.4	34.8	16400	3
N ₂	0.43	D	1.0	6.0	4.5	7-12	12.5	—	20300	4
N ₂	2.5	D	1.0	6.0	4.5	7	45.0	24.6	14500	4
He	0.9	C	0.25	3.5	3.5	20	24.3	—	61800	5
He	1.4	C	0.25	3.5	3.5	5-20	21.7	—	51500	5
He	2.8	C	0.25	3.5	3.5	5-20	24.6	—	41000	5
Ne	1.1	C	0.3	3.5	3.5	10	12.9	13.3	51400	6
A	0.57	D	1.0	6.0	4.5	5-10	10.5	—	17400	6
A	1.8	D	1.0	6.0	4.5	5	15.5	—	14200	6

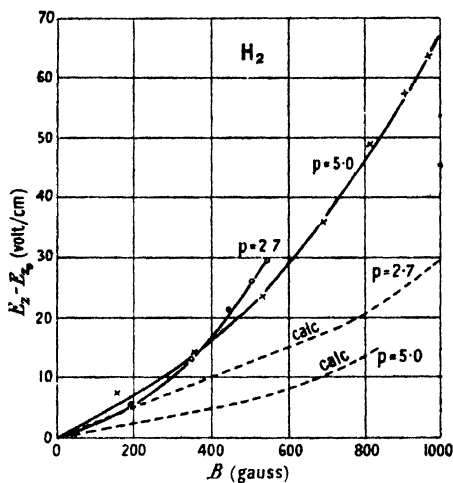


Figure 3. Increase in axial electric field as a function of magnetic field for hydrogen at different pressures and tube radii.

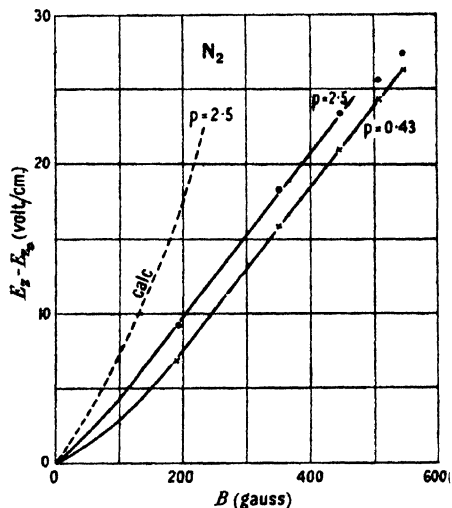


Figure 4. Increase in axial electric field as a function of magnetic field for nitrogen at different pressures.

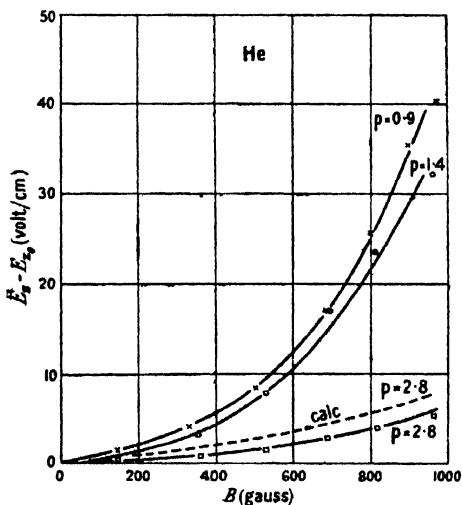


Figure 5. Increase in axial electric field as a function of magnetic field for helium at different pressures.

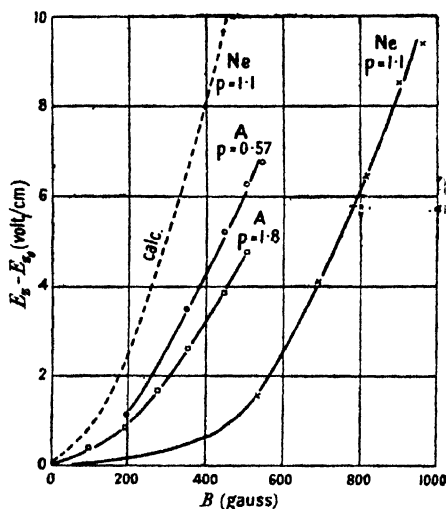


Figure 6. Increase in axial electric field as a function of magnetic field for neon and argon.

Full lines are experimental and broken lines are calculated from equations (9) and (10).

The voltage V between A and C (D) was measured with an ordinary voltmeter the resistance of which was 0.5×10^6 ohms. Thus the current through the instrument was less than 1 ma. and the probe voltage probably accurate to within a few volts. The electric field E_z is then approximately $E_z - E_{z0} = (V - V_0)/l$, where index zero refers to the magnetic field $B=0$. This method is not accurate but the measurements were made for the purpose of obtaining a first order check of the theory.

The tube data are given in the table and the relations between E_z and B are plotted in figures 3-6. The broken lines show the calculated relations. The values of κ used in the calculations are taken from von Engel and Steenbeck (1932, p. 187); these values are derived from Townsend's observations. κ varies very rapidly with the electron temperature, particularly in nitrogen; for helium κ is only known for elastic collisions, the occurrence of inelastic collisions increases its value. The value of κ has been obtained at three different electron temperatures by inserting in equation (9) the observed values of E_z and the calculated values of T at $B=0$; these values lie approximately on a straight line and give $\kappa = 3.4 \times 10^{-3}$ at $T = 40\,000^\circ \text{K.}$ and $\kappa = 9.3 \times 10^{-3}$ at $T = 60\,000^\circ \text{K.}$ Using these values in calculating the curves for $p = 2.8$ mm. gives the broken line in figure 5. The calculated curve for $p = 1.4$ mm. differs very little from that for $p = 2.8$ mm. The great difference between observed and calculated curves for the lower pressure is possibly due to the occurrence of strong high frequency oscillations which were observed.

For neon Druyvesteyn and Penning (1940) have given values of the electron energy losses in inelastic collisions from which κ can be calculated. Thus $\kappa = 0.058$ at $T = 60\,000^\circ \text{K.}$ and $\kappa = 0.092$ at $T = 80\,000^\circ \text{K.}$ Thus we obtain the broken line in figure 6.

In comparing the theoretical and the experimental results one must consider not only the simplifying assumptions made in the calculations and the uncertainty of the constants (especially κ) but also some disturbing effects in the measurements such as impurities in the gases. Only for the inert gases are the impurities supposed to have had negligible effects. The positive column was uniformly deflected towards the wall along the whole length of the tube. The anode glow was also influenced by the magnetic field and it is possible that the observed change in axial field was partly due to variations in the anode fall.

The measurements on nitrogen show a much higher value of axial field strength than calculation; this is probably due to traces of oxygen.

ACKNOWLEDGMENTS

I wish to express my great gratitude to Professor H. Alfvén, who has proposed this work to me, for his great interest and help in carrying out this investigation. I also wish to thank Dr. A. von Engel, Oxford University, for kindly reading the manuscript and for his valuable remarks.

REFERENCES

- DRUYVESTEYN, M. J., and PENNING, F. M., 1940, *Rev. Mod. Phys.*, **12**, 103.
 VON ENGEL, A., and STEENBECK, M., 1932 and 1934, *Elektrische Gasentladungen*, Bd. I and II (Berlin: Springer).
 TONKS, L., and ALLIS, W. P., 1937, *Phys. Rev.* **52**, 710.

Note on the Ionic Conductivity of Oxide-Coated Cathodes

By S. WAGENER

Thermionics Group of P.O. Engineering Research Station, Dollis Hill, N.W.2

MS. received 22 December 1948 and in amended form 12 April 1948

ABSTRACT. Ionic conductivity in an oxide-coated cathode is regarded as due to the mobility of oxygen ions. The magnitude of the conductivity can be derived from the phenomenon of current decay in a pulsed diode. Results are consistent with those obtained by other methods. Some consideration is given to the relation between ionic conductance and diffusion, and it is shown that the decay phenomenon results from the interaction of these phenomena in a thin layer on the outer surface of the coating.

§ 1. INTRODUCTION

VALUES of ionic conductivity of oxide-coated cathodes have recently been calculated by Sproull (1945) from decay times of pulsed emission currents. The coatings employed were of the mixed crystal type. Sproull assumed that the current decay occurred as a result of increased surface work function due to ionic conductance of barium ions away from the emitting surface. The values of ionic conductivity derived by Sproull, lying between 5% and 50% of the total conductivity, are rather high and do not correspond with values measured by other methods.

It is the object of this note to present an explanation of the decay phenomenon which differs materially from that of Sproull, and leads to results which are more consistent with known magnitudes of ionic conductivity.

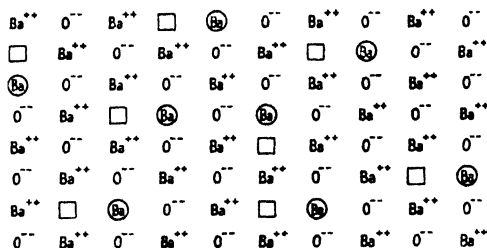
§ 2. ORIGIN OF IONIC CONDUCTANCE IN THE COATING

If the decay effect results from a temporal increase in surface work function during the current pulse, this increase may be regarded as due either to a flow of positive barium ions away from the surface, or to a flow of negative oxygen ions towards the surface. The second assumption is considered by the writer as the more probable and it will therefore be examined in some detail.

The oxide-coated cathode is now regarded as a reduction semiconductor containing an excess of barium atoms. The manner in which these excess atoms are built into the crystal lattice of the earth alkali oxide has been examined by Schottky (1935) and by Jost and Nehlep (1936). From their investigations of the monovalent NaCl lattice they concluded that the lattice defects of this compound are due to formation of vacancies. Similar conclusions were reached for the bivalent earth alkali oxides and in particular for BaO.

For example, in a pure BaO cathode, the crystal lattice of the unactivated coating will contain a certain number of vacancies at oxygen ion sites and a roughly equal number at barium ion sites. By activation an excess of barium is obtained, and this results from an increase in the number of vacancies in the lattice of oxygen ions together with a decrease in the number of vacancies of barium ions. A well activated cathode of BaO will therefore contain a large number of vacancies in the oxygen lattice and only a small number in the barium lattice. If the latter number is neglected, a cross section through the BaO lattice will follow the configuration

shown in the figure. The excess barium atoms, lying beside the vacancies, are indicated by circles in this figure.



BaO lattice in activated state.

The conductivities of the two kinds of ions depend on the number of vacancies belonging to each kind, and on their respective mobilities. Little knowledge exists of the relative magnitudes of these mobilities, but the number of vacancies appropriate to the oxygen ions has been shown to be much greater than the number appropriate to barium ions. There is a strong probability therefore that the part played by the oxygen ions in the total ionic conductivity is much greater than that of the barium ions. If, then, the decay phenomenon is to be explained by ionic conductivity, it is to the conductance of oxygen ions towards the surface rather than to the conductance of barium ions away from the surface that attention should be directed. Bearing this distinction in mind the problem can now be developed on lines similar to those employed by Sproull. To facilitate comparison with Sproull's paper, similar symbols have been used where possible and cross reference made to certain of Sproull's equations. The symbols are set out below.

N = number of oxygen ions per cm^2 of an atom layer of the BaO-crystal, N_c = number without vacancies, N_{min} = number for best activation, i = density of saturated emission current, e = electronic charge, t = time, κ_i = ionic conductivity, κ_e = electronic conductivity, D = diffusion constant, d = lattice spacing, Δ = thickness of layer in which ionic conductance and diffusion take place.

§ 3. CALCULATION OF THE DECAY OF EMISSION

When a current of density i traverses the coating from the surface to the core material, a number of oxygen ions are conducted to the atom layers near the surface. The flow of oxygen ions will reduce the number of oxygen vacancies near the surface, which in effect, results in an equal numerical reduction of excess barium there. The effect is analogous to the reduction of excess barium in the surface layer, such as occurs when oxygen ions impinge upon the surface from residual components of the valve atmosphere. In both cases the external work function increases, since this depends on the number of excess barium atoms in or near the surface. A decay of emission is therefore observed. The anomalous result of a decreased work function, observed by Sproull, may be due to the fact that he used the Richardson method which is not suitable for oxide-coated cathodes (Herrmann and Wagener 1944).

As $\kappa_e \gg \kappa_i$, the number of conducted ions N_i is given by

$$dN_i/dt = i\kappa_i/2e\kappa_e. \quad \dots\dots(1)$$

This flow of oxygen ions towards the surface sets up a concentration gradient through the coating and results in a diffusion flow of oxygen ions in the opposite direction. The number of these diffusing ions N_d is given by

$$dN_d/dt = -(N - N_0)D/\Delta d, \quad \dots\dots(2)$$

where the subscript 0 denotes quantities and currents at time $t = 0$. The effective number of ions, moving to the surface, is the vector sum of equations (1) and (2). Thus

$$\frac{dN}{dt} = \frac{\kappa_i i}{2e\kappa_e} - \frac{D}{\Delta d}(N - N_0), \quad \dots\dots(3)$$

which corresponds to equation (4) in Sproull's paper.

For evaluating equation (3) the dependence of emission current on the number of oxygen ions in the surface layer must be known. This dependence can be obtained from experiments of Becker (1929) who used the oxide coating as an anode and sent electrons, coming from a metal cathode outside, into the coating. This procedure gives exactly the opposite effect to that observed by Sproull. In Becker's experiments the ionic current, accompanying the electron current, flowed into the oxide coating from the surface towards the core. This resulted in a movement of oxygen ions from the surface into the interior and gave rise to the phenomenon of decreased external work function and increased emission. According to Becker the logarithm of the increase in emission is approximately a linear function of the number of oxygen ions removed. Let i_{N_c} and $i_{N_{min}}$ denote the saturated emission currents appropriate to surface states with N_c and N_{min} oxygen ions respectively. Then, according to Becker,

$$\ln(i/i_{N_c}) = b(N_c - N) \quad \text{and} \quad \ln(i_{N_{min}}/i_{N_c}) \simeq b(N_c - N_{min}). \quad \dots\dots(4)$$

Introducing the abbreviation

$$L \simeq \ln(i_{N_{min}}/i_{N_c}), \quad \dots\dots(5)$$

the ratio of the two equations in (4) gives

$$\ln(i/i_{N_c}) = L(N_c - N)/(N_c - N_{min}). \quad \dots\dots(6)$$

Equation (6) gives the required relation between emission current and density of oxygen ions in the atom layer at the surface. The equation differs from the corresponding equation (5) of Sproull in the value of the ratio on the right hand side. Sproull assumes the existence of a monatomic layer of barium on top of the coating and uses the ratio N/N_c (N denoting the density of barium instead of oxygen). The existence of such a monatomic layer, however, is very unlikely, if the composition of the activated oxide coating is as shown in the figure (cf. Herrmann and Wagener 1944).

If equation (6) is now rewritten for time $t = 0$, and the result subtracted from equation (4), then

$$\ln(i/i_0) = L(N_0 - N)/(N_c - N_{min}). \quad \dots\dots(7)$$

Differentiating and combining with (3) gives

$$\frac{1}{i} \frac{di}{dt} = \frac{-L}{N_c - N_{min}} \frac{dN}{dt} = \frac{DL}{\Delta d} \frac{N - N_0}{N_c - N_{min}} - \frac{\kappa_i Li}{2e\kappa_e(N_c - N_{min})}. \quad \dots\dots(8)$$

Combining (7) and (8) with

$$N_c - N_{\text{min}} = pN_c \quad \dots\dots(9)$$

results in the differential equation for the emission current

$$\frac{1}{i} \frac{di}{dt} = \frac{D}{\Delta d} \ln \left(\frac{i_0}{i} \right) - \frac{\kappa_1 L}{2e\kappa_0 p N_c} i. \quad \dots\dots(10)$$

§ 4. ASCERTAINING THE IONIC CONDUCTIVITY

The only difference between this and the corresponding equation (9) of Sproull is the appearance of the factor p . Solution of the equation leads accordingly to the same result as that of Sproull changed only by a constant factor. This factor is, however, decisive for the calculated magnitude of ionic conductivity. Thus, if $\kappa_1^{(\text{Ba})}$ is the conductivity calculated by Sproull for conductance by barium ions, and $\kappa_1^{(O)}$ is the conductivity for conductance by oxygen ions, then $\kappa_1^{(O)}/p = \kappa_1^{(\text{Ba})}/1$ or $\kappa_1^{(O)} = p\kappa_1^{(\text{Ba})}$.

It will be apparent from (9) that p is the relative difference between the number of oxygen ions in a complete atom layer and the corresponding number in a fully activated layer with a maximum number of vacancies. This difference is equal to the relative number of excess barium ions in the fully activated layer. According to most analyses the percentage of excess barium atoms in a fully activated cathode is of the order of 0.2% and therefore $p = 0.002$ (Clausing 1935, Isensee 1937). Hence the ionic conductivity, calculated for the conductance of oxygen ions will be only 2×10^{-3} times as large as the value calculated for barium ions by Sproull. From Sproull's values of $\kappa_1^{(\text{Ba})}$ between 5% and 50%, the derived values for $\kappa_1^{(O)}$ are therefore between 0.01% and 0.1%. These values correspond well with the value of 0.05% measured for BaO by Isensee (1937), but they are still somewhat higher than the value of 0.005% obtained for mixed crystals of BaO/SrO measured by Berdennikowa (1932) and 0.001% by Isensee. The earlier result of 0.5% by Becker (1929) is based on the assumption of a monatomic layer on top of the coating. Reconsideration of this result according to the concepts of surface described in this paper will give a value of the same order as those obtained by the other workers.

§ 5. APPLICATION OF THE RELATION BETWEEN IONIC CONDUCTIVITY AND DIFFUSION CONSTANT

Further consideration will now be given to the relation existing between ionic conductivity κ_1 and diffusion constant D . So far this relationship has been neglected, but its use leads to the removal of a further difficulty arising from Sproull's treatment. According to Wagner (1931)

$$D = kTu_1/v e, \quad \dots\dots(11)$$

where v = valency, k = Boltzmann's constant, T = absolute temperature, u_1 = ionic mobility. From the value of ionic mobility

$$u_1 = \kappa_1 / (v e N_0 / d), \quad \dots\dots(12)$$

is derived the relationship

$$D = 4.31 \times 10^{-5} T u_1 = 4.31 \times 10^{-2} (\kappa_1 d / 2 e N_0), \quad \dots\dots(13)$$

for a temperature of 1000° K.

Applying (10) for the final state of equilibrium in which $di/dt = 0$ and denoting the current flow during this state by i_∞ , an expression for D gives

$$D = \frac{\kappa_i L \Delta d}{2e\kappa_e p N_c} \times \frac{i_\infty}{\ln(i_0/i_\infty)}. \quad \dots\dots (14)$$

Using this value of D in (13) leads to an expression for the thickness of the layer of coating in which diffusion takes place. Thus

$$\Delta = 4.31 \times 10^{-2} \frac{\kappa_e p N_c}{L N_0 i_\infty} \ln(i_0/i_\infty). \quad \dots\dots (15)$$

As $N_0/N_c \simeq 1$, then for $\kappa_e = 5 \times 10^{-3}$, $L = \ln(i_{N_{\min}}/i_{N_c}) = 10$, $i_\infty = 0.5$ amp. and $\ln(i_0/i_\infty) = 2$, we find $\Delta = 17 \times 10^{-8}$ cm. With a lattice spacing for BaO of 2.76×10^{-8} cm., the derived value of Δ corresponds to about 6 atom layers.

The result indicates that the ionic conductance and diffusion which cause the decay of emission only take place in a thin layer on the surface of the coating. This is consistent with the fact, observed by Sproull, that the decay phenomenon is independent of the thickness of the coating. The absolute value obtained for Δ will not be considered here on account of the inaccuracy of the experimental values used for the calculation of Δ .

For further consideration of the problem it will become necessary to take into account the loss of oxygen ions from the outer surface by evaporation. Equation (10) might be adapted to cover this evaporation.

ACKNOWLEDGMENTS

The author is particularly indebted to Dr. G. H. Metson for his assistance with the English composition of this publication.

REFERENCE

- BECKER, I. A., 1929, *Phys. Rev.*, **34**, 1323.
 BERDENNIKOWA, T. P., 1932, *Phys. Z. Sowjet*, **2**, 77.
 CLAUSING, P., 1935, cf. DE BOER, J. H., *Electron Emission and Adsorption Phenomena* (Cambridge: University Press), p. 348.
 HERRMANN, G., and WAGENER, S., 1944, *Die Oxydkathode* (The Oxide Coated Cathode), Leipzig, Vol. II, pp. 151 and 168.
 ISENSEE, H., 1937, *Z. phys. Chem. B*, **35**, 309.
 JOST, W. and NEHLEP, G., 1936, *Z. phys. Chem. B*, **32**, 1.
 SCHOTTKY, W., 1935, *Z. phys. Chem. B*, **29**, 335.
 SPROULL, R. L., 1945, *Phys. Rev.*, **67**, 166.
 WAGNER, C., 1931, *Z. phys. Chem. B*, **11**, 139.

Emission of Secondary Electrons from Nickel and Molybdenum by Neutral Atoms of Mercury and Potassium

By R. M. CHAUDHRI AND A. W. KHAN

Physics Department, Muslim University, Aligarh, India

Communicated by M. L. Oliphant ; MS. received 14 June 1947, in amended form 29 April 1948

ABSTRACT. Investigations have been carried out on the emission of secondary electrons from Ni and Mo produced by high speed neutral atoms of mercury and potassium. The mercury particles were produced by neutralizing a homogeneous beam of accelerated ions from a hot-cathode mercury arc on the walls of a narrow nickel canal which they strike at grazing incidence. The neutral atoms of potassium were produced by passing a beam of fast ions of definite velocities through vapours of potassium where they are neutralized by electron exchange with neutral K atoms. It is observed that the secondary emission rises with the number and the energy of the particles and that the energy distribution among the secondary electrons is always Maxwellian, suggesting that the emission is probably due to intense local heating. The effective temperatures of the target have been calculated in typical conditions.

It is also shown that the targets scatter neutral atoms, dirty surfaces scatter more Hg atoms than clean ones but the reverse is the case with K particles.

§ 1. INTRODUCTION

THE emission of secondary electrons from metal surfaces by positive ions, both slow and fast, is well established but very few data are available for the secondary emission by neutral atoms. Webb (1924), Messenger (1926), Coulliette (1928) and Sonkin (1933) have investigated the secondary emission by very slow metastable atoms of mercury. Oliphant (1930) studied the energy distributions among the secondary electrons emitted from Ni and Mo by fast metastable atoms of helium.

The present work is devoted to the study of the secondary emission from Ni and Mo by high energy neutral atoms of mercury and potassium.

§ 2. APPARATUS

The apparatus and electrical connections are shown in figure 1. The filament F_1 , a thin tungsten spiral surrounded by a Ni spiral grid G , was held in position in the centre of a pyrex glass bulb A about 14 cm. in diameter. A Ni metal probe P , 4 cm. in diameter, with a circular hole 3 mm. wide at its centre and a Ni canal 5 cm. long and 3 mm. in diameter spot welded normal to its surface at the back, was sealed in a side tube in A opposite to F_1 . The Ni disc S_1 was 1.8 cm. in diameter and had a hole 3.5 mm. in diameter at its centre. The Ni cylinder S_2 , 4 cm. long and 1 cm. in diameter, had 4 mm. circular holes at the centre of its closed ends. The Ni cylinder C_2 , 3.5 cm. long and 2.5 cm. in diameter, was open at one end with a hole 5.5 mm. in diameter at the centre of its closed end and served to collect the secondary electrons emitted from the metal target, T , which was a circular disc 1.8 cm. in diameter. The Ni cylinder C_1 , 4.2 cm. long, 3 cm. in diameter, and open at one end had a circular hole 4.5 mm. in diameter at the centre of its other end and enclosed the cylinder C_2 to shield it from stray charges.

P_1 and P_2 were two rectangular parallel Ni plates each 2.5 cm. long and 1.5 cm. wide and were kept 5–6 mm. apart. The axes of the canal, the cylinders C_1 , C_2 , S_2 and the plates P_1 and P_2 were carefully adjusted to lie on the straight

line passing through the centre of S_1 . The target T was spot welded to a Ni tube C which, when desired, could be pulled back right out of the cylinder C_2 against the pressure of a tungsten spiral spring S . This is done by coiling the Ni flexible wire W round J by rotation of the ground-glass joint G_2 . The tungsten filament F_2 served to degas the target by electron bombardment. The whole apparatus was made of pyrex glass and all the metal parts were made of Ni or tungsten.

The apparatus was exhausted to a pressure below 10^{-6} mm. Hg by a two-stage all-steel mercury diffusion pump backed by a Cenco Hyvac pump and the mercury vapours were checked from entering the system by condensing them in the liquid air trap M . The experimental mercury in H was pure metal and was twice distilled under vacuum before use. The apparatus was tested to be vacuum tight and was baked at 500°C . for four hours before any observations were taken. The pumps were kept continuously running during the experiments and the gas pressure inside was read on a McLeod gauge.

The voltage for accelerating the particles was obtained from a D.C. motor generator set giving 0–5500 volts at 15 ma. and was read on electrostatic voltmeters.

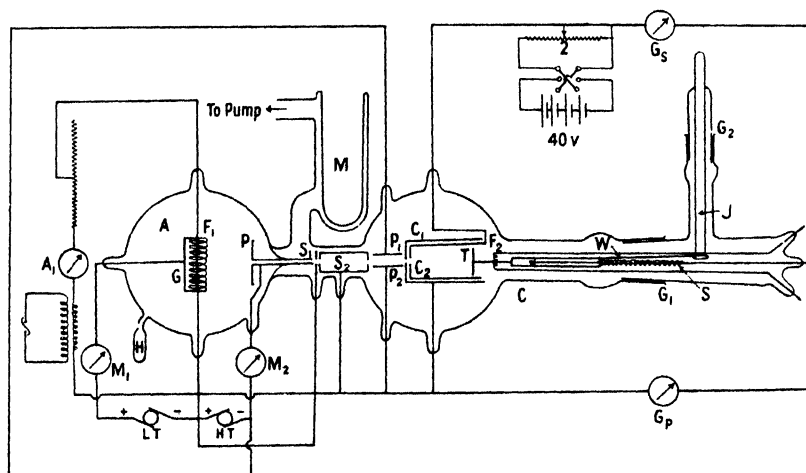


Figure 1.

The hot cathode mercury arc was maintained in A at 100–110 v. D.C. and the positive ions formed in the ionized gas were drawn to P by applying a negative potential to it with respect to the arc. Some of the positive ions arriving at P at normal incidence enter the canal and a fraction of them are neutralized by collision with its walls at grazing incidence without losing their energy, which is equal to the accelerating potential applied to the probe. The ions which pass right through the canal are stopped from reaching the target T and its surrounding system by a suitable reverse field applied between the canal and S_1 and S_2 and by the field between the parallel plates P_1 and P_2 . The accelerated neutral particles emerging out of the canal are, however, not influenced by these fields, but reach the target T and liberate therefrom secondary electrons which are collected on C_2 and measured by the galvanometer G_8 which has a current sensitivity of 10^{-10} amp/mm. Any positive ion current going to the target system is detected by the galvanometer G_p which is as sensitive as G_8 .

With electrical connections on as shown in figure 1, but with no heating current in F_1 , no deflection was observed in G_p or G_8 ; this indicates good electrical insulation of the whole system. On switching the mercury arc in A by heating the

filament F_1 no current whatsoever was detected by G_p but an appreciable electron current flowing from the target T to the collector C_2 was registered by G_a . That the emission was from the target was verified by pulling it right out of the collector; the current then fell to zero but recovered to its full value as soon as the target was brought in position inside C_2 . As no positive ion current was flowing to T or any of the cylinders C_1 or C_2 it was concluded that the observed secondary emission was either due to the radiations coming from the arc or was caused by the neutral atoms of mercury hitting the target or was due to both these effects. On switching off the positive ion accelerating voltage applied between F_1 and P , but keeping all other connections the same, the secondary current fell to zero, showing that the radiations from the arc were not causing this emission. It was therefore concluded that the neutral atoms of mercury formed in the canal and reaching the target were responsible for the observed secondary emission. The energy distribution among the secondary electrons was studied by varying the potential on the collector C_2 from $+40$ to -40 volts with respect to T with the help of the potentiometer Z and by observing the secondary current at each stage.

To obtain the neutral atoms of potassium the apparatus was slightly modified. The arc chamber A was replaced by a cylindrical pyrex glass bulb with a side tube containing pure potassium, distilled thrice in stages under vacuum. A Ni disc, 3 cm. in diameter and with a 3 mm. hole at its centre, took the place of the probe and was placed 2 mm. away from and opposite to the filament. Two Ni plates each 1.8 cm. in diameter and with a 3 mm. central hole were fixed behind the disc and replaced the canal, the distance between the disc and the last plate was about 5 cm. The rest of the apparatus was left unaltered. The tungsten filament was maintained at a dull red heat in potassium vapours kept at a temperature of $100-175^\circ\text{C}$. where the potassium atoms get ionized on striking against the hot tungsten surface and are accelerated to the Ni disc by applying a negative potential to it with respect to the filament. A beam of ions passes out of the hole in the disc and is further defined and collimated by the slit system. As the ions travel through the potassium vapour some of them get neutralized by electron exchange with neutral atoms but retain their initial kinetic energy. The residual positive ions are filtered out of the beam as in the case of Hg particles.

It is assumed that the neutral atoms of K and Hg produced as described above form a homogeneous beam and that all strike the target with the same energy, equal to the accelerating potential of the ions.

§ 3. RESULTS

Many experimental curves have been obtained under different conditions but only typical and essential ones are given here.

Figure 2*a* shows three retarding potential curves for the secondary electrons emitted from Ni by neutral atoms of mercury. Curve (1) was taken for an undegassed target with particles of energy of 2000 v. and with an ion current of 1.1 ma. to the probe while curves (2) and (3) represent results for a well degassed target for particles of energies of 2000 and 1500 v. respectively, the ion current to the probe in both cases being 3.2 ma. The straight lines in figure 2*b* show the relation between the logarithm of the secondary electron current I_s to C_2 and the retarding potential E_R applied on the collector for the corresponding curves in figure 2*a*.

Curves (1) and (2), figure 3*a*, give results with neutral mercury particles for dirty and clean Mo targets respectively. The energy of the particles in both cases was

2300 v. but the ion current to the probe was 1.4 ma. for the dirty and 3.2 ma. for the well degassed target. The semi-logarithmic curve for the well degassed Mo target is given in figure 3b.

Curves (1) and (2), figure 4a, show results obtained with neutral atoms of potassium of 1500 v. energy for a dirty and a well degassed Mo target respectively. The ion current to the disc was kept the same in both cases. The corresponding $\ln i - E_R$ curves are given in figure 4b. Similar curves were obtained for a Ni surface.

It is seen from all the retarding potential curves in figures 2, 3 and 4 that with

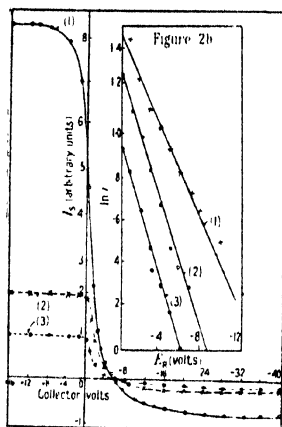


Figure 2a. Results with neutral Hg atoms on Ni.

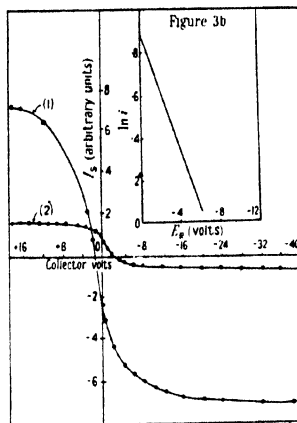


Figure 3a. Results with neutral Hg atoms on Mo.

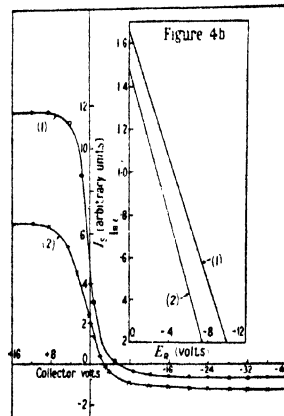


Figure 4a. Results with neutral K atoms on Mo.

- (1) Target=Dirty Ni.
Probe current=1.1 ma.
Ion accelerating volts=2000.
- (2) Target=Degassed Ni.
Probe current=3.2 ma.
Ion accelerating volts=2000.
- (3) Target=Degassed Ni.
Probe current=3.2 ma.
Ion accelerating volts=1500.

- (1) Target=Dirty Mo.
Probe current=1.4 ma.
Ion accelerating volts=2300.
- (2) Target=Degassed Mo.
Probe current=3.2 ma.
Ion accelerating volts=2300.

- (1) Target=Dirty Mo.
Ion current to disc=1.2 ma.
Ion accelerating volts=1500.
- (2) Target=Degassed Mo.
Ion current to disc=1.2 ma.
Ion accelerating volts=1500.

dirty targets the secondary current to C_2 is saturated at collector potentials 4–16 v. positive with respect to T and at lower voltages with degassed surfaces. In the case of clean Ni targets the collector is positive with respect to T only by a fraction of a volt when saturation occurs. Below saturation potential the current to C_2 decreases with the collector voltage. The secondary electron current reverses and flows from the collector to the target T when the retarding potential on C_2 exceeds a limit. The reverse current increases first as the negative potential on C_2 is raised but soon settles down to a saturated value. Some of the neutral particles incident on the target are reflected back from it and hit the walls of the cylinder C_2 , where they liberate secondary electrons which go to the target when the latter is positive with respect to the collector. At a certain negative potential on C_2 this current balances the secondary emission current from T but is greater than this at higher retarding voltages. The reverse current will show saturation when the collector voltage is big enough to stop all the secondary electrons of the target from reaching C_2 . The secondary electrons from the target are found to possess all

energies in the range 0–20 v. and the energy distribution among them is Maxwellian. Assumption of a similar energy distribution among the secondary electrons liberated from C_2 will explain why the secondary current to the collector does not attain saturation until the positive potential of C_2 with respect to T is greater than a certain value. In the absence of the secondary emission from the walls of C_2 the secondary current to it would show saturation with collector potential equal to or only slightly higher than that of T.

The reverse current is much larger with dirty targets than with clean degassed surfaces. In the case of a dirty Mo target the reverse current is of the same order as the secondary electron current flowing from T to C_2 . This decreases appreciably when the target is well degassed, but even then it is much larger than in the case of a clean Ni surface under almost the same conditions, as is seen by comparing curves (2) in figures 2 and 3. It is concluded that the neutral particles of Hg are scattered more from dirty surfaces than from clean ones and that a Mo surface scatters more than Ni. Results with K atoms on Mo and Ni show, however, that more K atoms are scattered from a clean surface than from a dirty target.

The variation of the secondary emission from degassed Ni with the energy of the particles in the range of 1500–2700 v. for mercury and of 1500–2500 v. for potassium is shown in curves (1) and (2) of figure 5. The intensity of the neutral

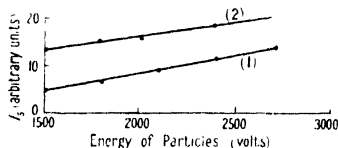


Figure 5.

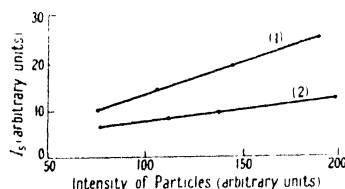


Figure 6.

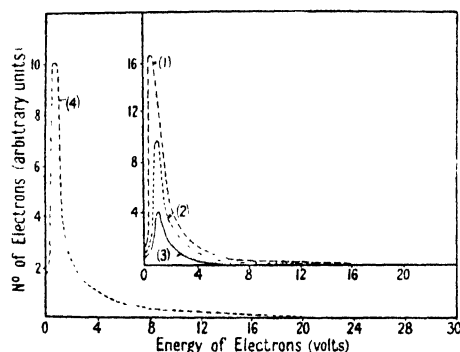


Figure 7.

- (1). Undegassed Ni target. Energy of particles = 2000 volts.
- (2). Degassed Ni target. Energy of particles = 2000 volts.
- (3). Degassed Ni target. Energy of particles = 1500 volts.
- (4). Degassed Mo target. Energy of particles = 2300 volts.

particles in each case was maintained constant throughout the range by drawing a fixed positive ion current to the target at different energies and then filtering out the positives from the mixed beams. Curves (1) and (2) of figure 6 show the change in the secondary emission with the intensity of mercury particles of energy of 2000 v. from clean Ni and Mo targets respectively. Similar results were obtained with K particles on Ni targets.

The energy distribution among the secondary electrons was obtained for different cases from the corresponding retarding potential curves and is typified

by the curves given in figure 7 which are for Mo and Ni targets hit by neutral particles of Hg of varying energies. The distribution in all cases is found to be Maxwellian, the maximum of the curves occurring near zero energy. The position of the peak remains unaltered with variation of the energy of the particles and the condition of the target, but the maximum energy of the secondary electrons increases with the energy of the neutrals.

The secondary emission is always greater from a dirty target than from a clean one and rises almost in proportion to the energy of the particles over the range studied and increases linearly with their intensity as shown in figures 5 and 6.

The secondary electron current in these experiments is of the order of 10^{-8} amp.

The Maxwellian energy distribution among the secondary electrons suggests that the emission is caused by intense local heating of the target spot by the incident neutral particles. The effective temperatures of the target obtained from the semi-logarithmic curves are given below for typical conditions.

Mercury atoms of energy 1500–2500 v.

Undegassed nickel	40 000–45 000° K.
Degassed nickel	29 000–34 000° K.
Undegassed molybdenum	40 000–45 000° K.
Degassed molybdenum	30 000–33 000° K.

Potassium atoms of energy 1200 v.

Undegassed molybdenum	35 000° K.
Degassed molybdenum	29 000° K.

ACKNOWLEDGMENTS

Our thanks are due to the Muslim University for providing facilities for carrying out this work.

We wish also to express our sincere thanks to Professor M. L. Oliphant for his helpful advice and encouragement.

REFERENCES

- COULLIETTE, H. J., 1928, *Phys. Rev.*, **32**, 636.
 MESSENGER, H. A., 1926, *Phys. Rev.*, **28**, 962.
 OLIPHANT, M. L., 1930, *Proc. Roy. Soc. A*, **127**, 373.
 SONKIN, S., 1933, *Phys. Rev.*, **43**, 788.
 WEBB, H. W., 1924, *Phys. Rev.*, **24**, 113.

The Scattering of Fast β -particles through Large Angles by Nitrogen Nuclei

By F. C. CHAMPION AND R. R. ROY

Wheatstone Physics Laboratory, King's College, University of London

Communicated by J. T. Randall; MS. received 8 June 1948, read 5 November 1948

ABSTRACT. A combined arrangement of counters and a cloud chamber is used to examine the elastic large-scale scattering of β -particles of approximately 1 mev. energy by nitrogen nuclei. In a total effective track length of nearly 100 km., 131 photographs were obtained in which the scattering angle exceeded 85° . Considering the present results on large-angle scattering together with earlier results on the scattering at smaller angles, the agreement between experiment and Mott's theory is good over the entire angular range 20 – 180° .

§ 1. INTRODUCTION

IT is now generally accepted that the scattering through small angles of β -particles of energy up to a few mev., by light nuclei, is in agreement with the theory as initially developed by Mott (1929). Most of the earlier experiments were made by the expansion chamber method (Champion 1936); more recent experiments using electrical methods of recording have confirmed and extended these results (Van de Graaff *et al.* 1946). No extensive experiments, however, have hitherto been performed to examine the scattering at angles greater than 90° . At high angles of scattering, the small intensity compared with the general background makes measurement by electron gun methods very difficult. If the expansion chamber method is used in the ordinary way, only a fraction of one per cent of the photographs show large angle deflections. This results in a waste of photographic material and involves tedious and prolonged scrutiny of an enormous number of recordings in the search for rare events. It was therefore decided to fit a standard cloud chamber with counter control in such a way that photographs were obtained only when β -particles suffered deflections greater than about 85° .

§ 2. EXPERIMENTAL ARRANGEMENT

In figure 1 is shown a diagram of the arrangement of the expansion chamber and the counters. Three holes were bored in the top glass plate of the chamber so that the centres of the holes lay on the quadrant of a circle whose centre was at the mid-point of the plate. Into these three holes were inserted three aluminium counters C_1 , C_2 and C_3 of length 6 cm. and diameter 2 cm., so that they extended into the full depth of the expansion chamber. The procedure consisted in making expansions at regular intervals of about one minute in the usual way, with the photographic plate permanently exposed. Illumination occurred and a record was obtained only when any one of the counters was discharged by the entry of a β -particle which had been scattered through a large angle.

Accidental discharges of the counters by cosmic rays and other extraneous effects could not be entirely prevented, but illumination during periods other than the sensitive time of the chamber was eliminated by a further control mechanism shown in figure 2. The piston release of the expansion chamber closed K_1 and K_2 at the beginning of an expansion, thus short-circuiting the magnet M_1 and hence releasing the pendulum P_1 .

The short time interval between the release of this pendulum and its impact on the key K_3 was the period of sensitivity for β -particle recording. During that time, if the Geiger-counter recorder circuit is closed by the passage of a β -particle through one of the counters in the chamber, the magnet M_2 is short-circuited and the consequent release of the pendulum P_2 operates the flash mechanism for photographing the tracks. Impact of P_1 on K_3 breaks the circuit which, if complete, would allow short-circuit of M_2 should the counter be discharged. Hence at all times other than the short interval between the release of P_1 and its impact on K_3 , the flash mechanism is inoperative. In this way, unwanted photographs

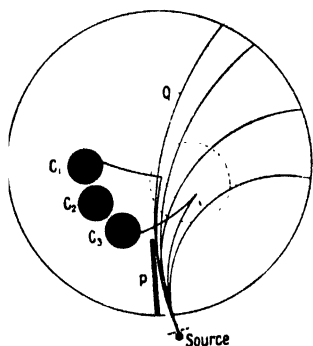


Figure 1.

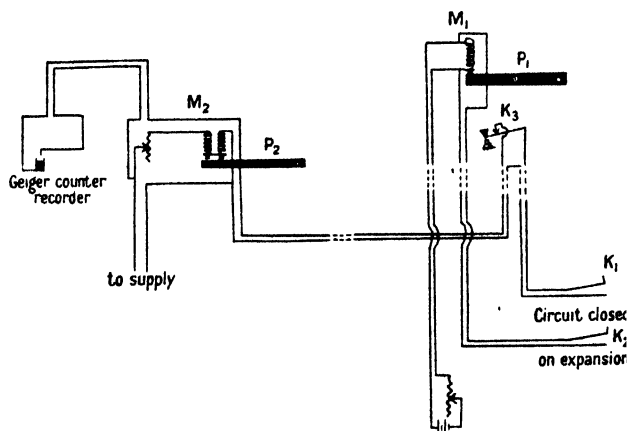


Figure 2.

produced by discharges other than those resulting from a scattered β -particle were reduced to one in forty. Some 1500 expansions were made for every photograph obtained. As shown in figure 1, a lead plate P , which projected 4 cm. into the chamber on one side of the window which admitted the β -particles from a RaE source (placed just outside and fitted with a shutter synchronized with the expansion), prevented the discharge of the counters by β -particles other than those scattered through large angles. A uniform magnetic field was provided by two Helmholtz coils in the usual way and final measurements were made on the images of the tracks by the standard method of reprojection.

§ 3. RESULTS

In a total effective track length of nearly 100 km., over 100 photographs were obtained in which β -particles of 0.5 to 1 mev. energy were scattered by nitrogen nuclei through angles greater than 85° . The improvement in the present arrangement may be judged by considering that in previous work, the tedious examination of 1000 photographs yielded under 1 km. of track and showed only four such large angle collisions. The distribution of the intensity of scattering with angle is

shown in the second column of the table. In the third column, the observed values are corrected by a factor which determines the relative chance of recording particles in the given angular range. The last column contains the intensity to be expected theoretically according to Mott's formula for light elements applied to nitrogen.

Angular interval	Observed	Obs. (corrected)	Theoretical
85–100°	74	88	55
100–120°	34	34	37
120–180°	23	23	26
85–180°	131	145	118

§ 4. DISCUSSION

Examination of the table shows that taking the angular distribution as a whole between 85° and 180°, the calculated and the measured intensities are in fair agreement. However, although the present experimental arrangement clearly has considerable advantages over older methods without counter control, certain important precautions are necessary in selecting the data and in interpreting the results. For example, a particle scattered at some point Q in figure 1 would not be recorded if the angle of scattering were 90° since it would not pass through the counters. A similar particle would, however, be registered if it were scattered through a considerably larger angle. In order to make this selectivity of the apparatus as small and as definite as possible, the following criterion was introduced. The scattered particles recorded in column 2 were all produced in a track length of about 4 cm. located in the region enclosed in the dotted circle in figure 1. If the measurements are restricted to this definite region, then by empirical construction, involving the drawing of a large number of typical tracks, an estimate may be made of the degree of selectivity introduced.

The radius of curvature of the tracks varied from about 10 to 14 cm. over the range of the β -ray spectrum considered. On a full scale diagram typical tracks are drawn and these are marked along their length at regular intervals. From these marks branch tracks are drawn representing the bending of the tracks through various angles. Suppose that beyond a certain mark, all tracks bent through less than 90°, for example, no longer meet any one of the counters and are therefore lost. If the total track length from which deflections are measured is y cm. and if the particles bent through the angular range 85–95° are not recorded along a length of more than x cm., the correction factor for these deflections compared with other fully recorded deflections at larger angles is clearly y/x .

Restriction to a definite region also allows the calculation of the effective solid angle subtended by the counters, into which the scattering is recorded. A further point is that since the counters were not touching each other, a certain number of particles passed in the gaps between the counters and were not recorded. These corrections were small for the largest angles but rose to nearly 20% for angles in the region of 90°. It is these factors which account for the difference between the second and third columns of the table.

Counters similar to those used in this experiment had been used in previous work (Widdowson and Champion 1938) and the indications are that for β -particles of the energy and intensity under consideration here, they are 100% efficient. Indeed, the number of particles recorded exceeded somewhat that to be expected theoretically at the smaller angles. It is very unlikely that the discrepancy is real.

It is far more likely that some tracks whose angle of scattering is somewhat less than 85° have been recorded. Since the angle cannot be measured accurately to within a few degrees and the general tendency of the observer is to reject as little data as possible, the fact that the scattering intensity rises very rapidly with decreasing angle could account for the apparent excess at the rather uncertain lower angular limit.

No inelastic collisions of fast β -particles with nitrogen nuclei were recorded with this arrangement, which however, was unsuited to a satisfactory examination of inelastic scattering for two reasons. First, the remaining energy of the electron after collision might be insufficient to penetrate the counter wall, and secondly, the radius of curvature of the track would be reduced after inelastic collision so that it might not traverse the region occupied by the counters. Both these difficulties could be overcome to some extent by using several thin-walled counters distributed over a greater length of arc in the chamber than was used in the present arrangement. However, since such inelastic collisions in nitrogen have been found to be very infrequent it is doubtful if such a straightforward extension of our present experiment would add very much to our knowledge.

Combining these experiments on the large-angle elastic scattering of β -particles of energy up to 1 mev. by nitrogen nuclei with our previous results on the scattering at smaller angles, we can state that Mott's formula is in good agreement with experiment at all angles of scattering.

REFERENCES

- CHAMPION, F. C., 1936, *Proc. Roy. Soc. A*, **153**, 353.
MOTT, N. F., 1929, *Proc. Roy. Soc. A*, **124**, 426.
VAN DE GRAAFF, R. J., BUECHNER, W. W., and FESHBACH, H., 1946, *Phys. Rev.*, **69**, 452.
WIDDOWSON, E. E., and CHAMPION, F. C., 1938, *Proc. Phys. Soc.*, **50**, 185.

The Penetrating Particles in Cosmic-ray Showers: II. The Lightly-ionizing Penetrating Particles in Penetrating Showers

BY G. D. ROCHESTER AND C. C. BUTLER

The Physical Laboratories, The University, Manchester

MS. received 26 July 1948

ABSTRACT. The momentum spectrum of the penetrating particles in a small sample of penetrating showers has been measured. Most of the particles are positive and have momenta of the order 10^9 ev/c. From a study of the interaction of these particles with nucleons in a lead plate it is concluded that whilst a small number may be μ -mesons most are protons, π -mesons, or heavier mesons.

§ 1. INTRODUCTION

AN account is given of the lightly-ionizing, penetrating particles occurring in a small group of penetrating showers selected from a larger group of cosmic-ray showers. For the purpose of this analysis a penetrating shower is defined as a shower containing two or more ionizing penetrating particles. Penetrating showers containing single penetrating particles in the cloud chamber have been excluded because of the difficulty of distinguishing them from knock-on showers. A selection has been made, therefore, which is much more rigid than in counter work on penetrating showers.

Broadbent and Jánossy (1947) have suggested that penetrating showers are of two types, namely, local and extensive. A local penetrating shower is one created in the layer of absorber immediately above the counter set while an extensive penetrating shower is one associated with an air shower. It seems probable that the local showers are produced by fast nucleons (Jánossy and Rochester 1943), and it is to this type of shower that the results in this paper mainly refer.

The nature of the penetrating particles in penetrating showers is not yet known but there is much evidence that the local showers contain mesons and protons (Hazen 1944, Daudin 1944, W. M. Powell 1946, Rochester 1946, Rochester, Butler and Runcorn 1947, Rochester and Butler 1947, Fretter 1948). The results presented in this paper, while supporting this conclusion, indicate that only a small fraction of the mesons are μ -mesons.

§ 2. THE EXPERIMENTAL ARRANGEMENT AND THE MEASUREMENT OF MOMENTUM

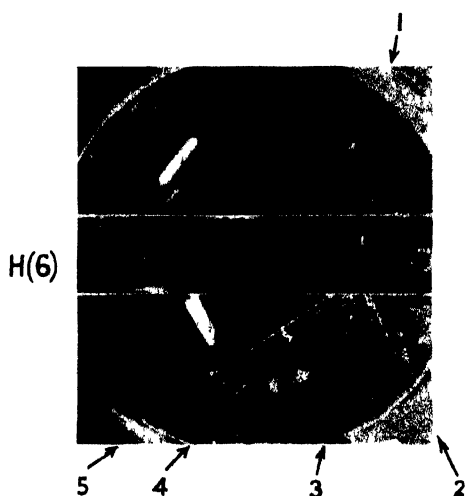
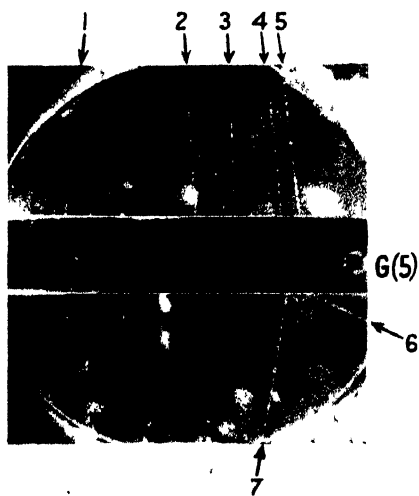
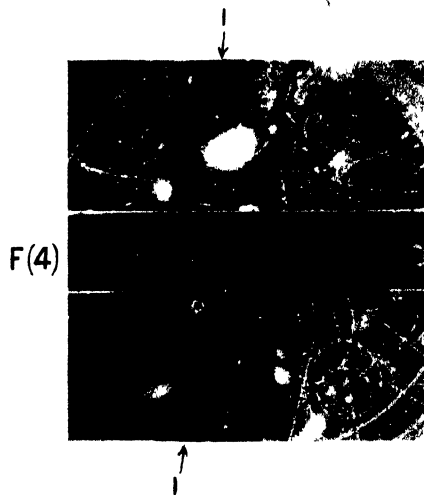
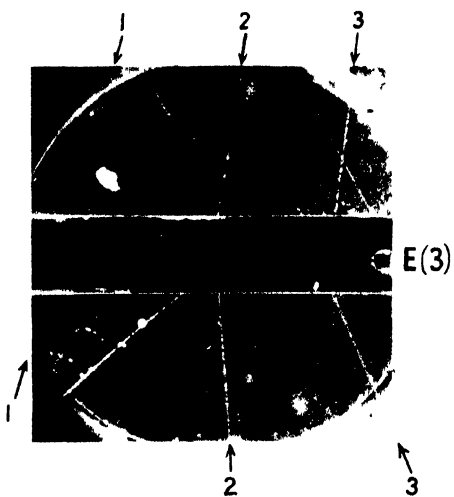
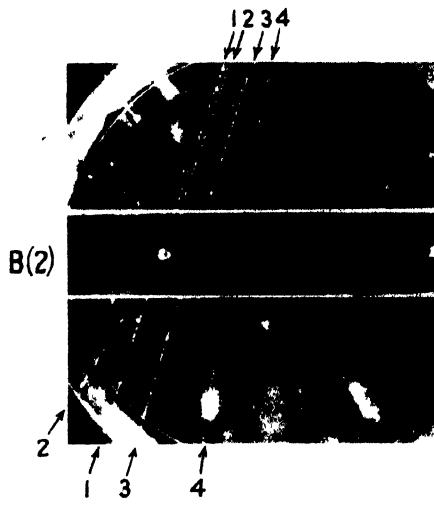
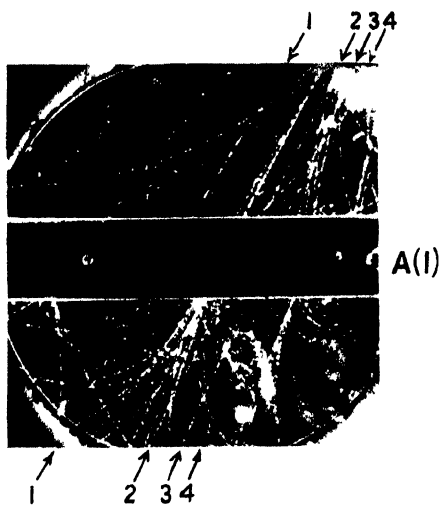
The experimental arrangement was described in an earlier paper (Rochester and Butler 1948, to be referred to as I). All the showers described in the present paper were taken with counter arrangement P, figure 2 of paper I. The thickness of lead in contact with the wall of the chamber and the total thickness of lead above the chamber are given in table 1. There was no lead below the chamber.

As described in I the curvatures of the tracks were measured on the Blackett curvature-compensating machine. Showers were selected with tracks free from

Table 1. Penetrating particles in penetrating showers

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
A(1)	1.8	16.8	7200	1	Positive	1.3	0.1	1.4
				2	Positive	1.1	0.1	1.7
				3	Positive	1.4	0.5	1.3
				4	Positive	0.65	0.8	2.9
B(2)	1.8	16.8	7250	1	Positive	3.3	0.6	0.6
				2	Positive	0.91	12.8	2.1
				3	Positive	1.0	0.8	1.9
				4	Too faint to measure			
C	1.8	16.8	7100	1	Positive	0.66	0.5	2.9
				2	Negative	0.85	2.2	2.2
D	1.8	6.8	7200	1	Positive	1.2	2.4	1.6
				2	Positive	1.1	0.0	1.7
				3	Positive	1.0	0.0	1.9
				4	Positive	0.77	2.4	2.4
E(3)	5.0	10.0	6900	1	Too faint to measure			
				2	Negative	1.1	12.0	1.7
				3	Positive (above)	4.5	28.0	—
					(below)	0.3	—	—
F(4)	5.0	10.0	7300	1	Positive	0.63	3.0	3.0

- (1) Shower (No. of photograph on Plate). (2) Lead in contact with wall of chamber (cm.).
 (3) Total thickness of lead above chamber (cm.) (4) Magnetic field (gauss).
 (5) Number of particle. (6) Sign of particle. (7) Momentum ($\times 10^9$ ev/c.).
 (8) Angle of scatter (deg.) observed. (9) Angle of scatter (deg.) calculated.



PLATE

For all photographs except (3) a positive particle coming down is curved in a clockwise direction.

Photograph (1): A complex penetrating shower consisting of four lightly ionizing penetrating particles and a large electronic shower. The penetrating particles are indicated by numbers. They are positively charged and have momenta ranging from 1.4 to 0.65×10^9 ev/c.

$H = 7200$ gauss.

Photograph (2): An example of a simple type of local penetrating shower consisting of four positive penetrating particles whose momenta range from 3.3 to 0.91×10^9 ev/c. Particle 2 is anomalously scattered through 12.8° in the lead plate.

$H = 7250$ gauss.

Photograph (3): A penetrating shower consisting of three penetrating particles 1, 2 and 3. Particle 2 is a negative particle of momentum 1.1×10^9 ev/c. and is anomalously scattered through 12.0° in the lead plate. Particle 3 is positive with a momentum of 4.5×10^8 ev/c. above the plate and 3.0×10^8 ev/c. below the plate and is scattered through 28.0° . The particle below the plate is a proton. The heavily ionizing particle which appears to come from the same region in the lead plate as the lower part of 2 is actually in a plane 1.8 cm. behind it. Thus if the heavily ionizing particle is connected with particle 2 it must be via an intermediate link. The heavily ionizing particle is positive and has a momentum of 1.6×10^8 ev/c. A proton of this momentum ionizes 15° minimum whereas the ionization is estimated as 7° minimum. The difference may be due to fluctuation or indicate a particle of intermediate mass.

$H = 6900$ gauss.

Photograph (4): A complex penetrating shower associated with an extensive shower. Only particle 1 is clearly penetrating but at least three other particles in the wide core seem to pass through the plate without multiplication. The shower is coming forward in the chamber at a rather steep angle. Accurate measurement of the tracks is difficult because of the confusion and the rather low technical quality. Most of the particles seem to be positive and some are lightly ionizing with momenta approximately 5×10^8 ev/c. suggesting that they are not protons. The wide-angle pair of tracks at the lower right-hand side of the photograph seem to be those of protons.

$H = 7300$ gauss.

Photograph (5): An unusual shower consisting of several high-energy particles which stop in the lead plate without producing visible particles. All the particles except 3 are well in the illuminated region of the chamber and nearly all have momenta much above 10^9 ev/c.

$H = 7100$ gauss.

Photograph (6): An example of a high-energy star induced by particle 1. Particles 4 and 5 have momenta of approximately 10^9 ev/c. and are positive and negative respectively. This photograph would appear to indicate the creation of mesons or protons.

$H = 6600$ gauss.

obvious distortion and no track was measured unless its length in the chamber was greater than 6 cm. A check on the performance of the cloud chamber was made in two ways, firstly, by measuring up 87 meson tracks photographed without magnetic field, and secondly, by measuring up 90 meson tracks in the magnetic field. All these tracks were taken under chamber conditions identical with the showers. The lead above the chamber was, however, removed. It was found that a fairly good Gaussian curve could be fitted by the method of least squares to the no-field curvature measurements. The half-width of the Gaussian corresponded to a momentum of 8.4×10^9 ev/c. in a magnetic field of 7500 gauss, which may be taken as the maximum detectable momentum (Blackett 1937).

The meson tracks in the magnetic field provided an effective check of the performance of the cloud chamber under conditions identical with those for penetrating showers. The resulting meson spectrum agreed closely with the data given by Wilson (1946). It may therefore be assumed that the performance of the cloud chamber was satisfactory.

§ 3. THE PENETRATING PARTICLES IN PENETRATING SHOWERS

(i) *Photographs and classification of the showers*

Using the criteria outlined in §§ 1 and 2, six penetrating showers containing sixteen penetrating particles have been selected for measurement and the results are given in table 1. The number in brackets after the letters in column (1) of table 1 refer to the photographs on the plate. A photograph of shower D has already been published (shower 2 of Rochester and Butler 1947). With the photograph published by Rochester, Butler and Runcorn (1947) these photographs present the main types of penetrating shower in a magnetic field. It is seen that a broad classification is possible in that some penetrating showers consist only of a few ionizing penetrating particles as, for example, showers B, D and E, while others include a considerable electronic component, as for example, showers A and F. This classification is clearly shown in the investigation made by Rochester (1946) and by Bridge, Hazen and Rossi (1948). Showers of the first type are presumably local penetrating showers, for in each case the penetrating particles can be traced back to a common point of origin in the absorber. Showers of the second type may be associated with extensive showers. Shower F was connected with an extensive air shower but it is not possible to classify shower A because the counter extensive tray was not in operation when this photograph was taken.

(ii) *Positive excess and spectrum*

The data show several interesting features, two of the most striking being the large positive excess and the relative abundance of particles with momenta

Table 2. Spectrum of penetrating particles

Momentum range ($\times 10^9$ ev/c.)	1-5	6-10	11-20	21- ∞
No. of penetrating particles.	0	8 ± 3	6 ± 2	2 ± 1

below 2×10^9 ev/c. It will be noticed that of the 16 penetrating particles, 14 are positive and only 2 negative. This large positive excess suggests that some of the penetrating particles are protons. However, with such a small sample

the result should be taken with caution. Nevertheless it is interesting to see how many of the particles are probably *not* protons. We may assume that the negative particles are not protons and if we add the particles which are still lightly ionizing when their momenta are less than 0.66×10^9 ev/c.—a proton of this momentum would ionize twice minimum—we find five out of the sixteen particles.

(iii) *Anomalous scattering*

Another striking feature is the anomalous scattering of three of the penetrating particles in the lead plate. These particles are: particle 2 of shower B, which is scattered through 12.8° , and particles 2 and 3 of shower E, which are scattered through 12.0° and 28.0° respectively. The probable angle of scatter for any one of these particles, calculated from Williams' formula, is not greater than 2° . There is little doubt that this result is significant especially when taken in conjunction with previous work on penetrating showers. Thus, Rochester (1946) found two cases of particles scattered through 10.2° and 18.0° respectively out of 32 penetrating particles in penetrating showers. Although the momenta of the particles were not known, the type of shower was so similar that it may be assumed that these were also cases of anomalous scattering. In all, five cases of scattering have been found in 48 penetrating particles or, in terms of the thickness of lead traversed, one particle per 25 cm. lead. This value is in sharp contrast with the scattering of ordinary mesons. Thus Wilson (1940) found one case of anomalous scattering per 40 m. of lead traversed, and Shutt (1946) and Code (1941) have obtained similar results. Where the sign could be determined almost all of the anomalously scattered particles were positive. It has therefore been suggested that the observed cases of anomalous scattering are due to protons and that the observed scattering cross-section does not represent the true cross-section for mesons. The results presented here support this conclusion; indeed, particle 3 of shower E can be unambiguously identified below the plate as a proton. It is not, however, valid to assume that the particles above and below the plate are necessarily the same, nor even that they are all protons. They could, for example, be protons or π -mesons which produce stars in the plate with secondary particles of such short ranges that all are absorbed except one. It is improbable that the scatterings represent examples of the spontaneous disintegration observed by Rochester and Butler (1947) in shower D, because only one such event has been found in the gas of the chamber, whereas at least five times as many events should have been observed in the gas as in the lead plate. Thus, it must be concluded that the scatterings are due to a collision process and not to any type of spontaneous process.

The minimum momenta of a meson of mass $200 m_e$ and a proton to penetrate the lead plate are 1.4×10^8 ev/c. and 5.7×10^8 ev/c. respectively. A proton of this momentum would ionize 2.4 times minimum.

§4. SHOWERS G AND H

In this paragraph a description will be given of two unusual showers which are shown in photographs 5 and 6 of the plate. The data for these showers are given in table 3. The positions of the ends of the tracks with respect to the front of the chamber, which is 9 cm. in depth, are given in columns (8) and (9), so that some idea of the space-orientation of the tracks can be obtained. The lengths of the tracks vary from 6 cm. to 12 cm. in the chamber.

(i) Shower G

This shower was connected with an extensive air shower and like showers F and E consists mainly of a collimated group of positive particles. The tracks are, however, not quite parallel, but project back to a common point of origin in the lead absorber above the counter set (see figure 2, paper I). It is clear that these particles are not electrons for most have momenta greater than 10^9 ev/c. and yet do not produce showers in the lead absorbers either above or in the cloud chamber. According to Chakrabarty (1942) an electron of momentum 10^9 ev/c. produces on the average a shower of 10 particles in a lead plate of 3 cm. thickness and an electron of 5×10^9 ev/c. a shower of 50 particles. The particles seem to be similar to the penetrating particles discussed in §3. Particle 5 is scattered in the plate and emerges as particle 6; particle 7 is probably associated with the same collision process. The other particles are remarkable in that they are apparently completely absorbed in the lead plate without producing other charged particles. Although some are moving slightly backwards, their inclinations

Table 3. Showers G and H

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
G(5)	5.0	10.0	7100	1	Positive	0.75	4.9	3.8
				2	Positive	7.0	4.0	2.9
				3	Positive	8.5	6.0	4.0
				4	Positive	2.3	2.8	2.8
				5	Positive	1.7	4.0	4.0
				6	—	—	3.5	3.5
				7	—	—	4.5	4.5
H(6)	5.0	10.0	6600	1	Positive	5.2	3.3	3.3
				2	—	—	2.8	2.8
				3	—	—	2.8	2.8
				4	Positive	1.5	2.1	1.3
				5	Negative	1.1	4.0	6.2

- (1) Shower (No. of photograph on Plate.) (2) Lead in contact with wall of chamber (cm.).
 (3) Total thickness of lead above chamber (cm.). (4) Magnetic field (gauss).
 (5) Number of particle. (6) Sign of particle. (7) Momentum ($\times 10^9$ ev/c.). (8)
 Position of track at plate (cm.) (9) Position of other end of track (cm.).

and positions at the plate do not suggest that they would have gone out of the illuminated region. Just as in the case of the large-angle scattering it is not known what type of collision process will account for this phenomenon. A large star, could however, be produced in the plate and no visible particle appear.

This shower and several of the others, for example, showers B, E and F also show another very interesting phenomenon, namely, tracks which are almost parallel in the cloud chamber. In at least three of the cases the tracks project rather accurately back to the lead absorber which is 75 cm. above the centre of the cloud chamber. It is thus quite possible that the showers are originally star-like and that geometrical factors cause the selection of an almost parallel core of particles in the cloud chamber. Low-energy penetrating particles, if not lost by their emission at wide angles, may be lost by absorption in the lead immediately above the cloud chamber. Thus the showers presented here are quite consistent with the star-like penetrating showers found by Hazen, Fretter

and others. Moreover, they are also consistent with production by incident particles whose ranges are from 5 to 10 cm. of lead. In order to account for showers like E, F and G, however, where particles emerge from a block of lead 5 cm. in thickness and are then absorbed in a further 3 cm. lead plate, we require that many of the secondary particles are, like the primary particles, catastrophically absorbed with a range of from 5 to 10 cm. lead.

(ii) *Shower H*

This shower is of especial interest because it is an explosive-type shower in which high-energy particles are emitted at large angles to the direction of the incident particle. It will be seen from table 3 that the incident particle 1 has a very high energy. The particles which emerge below the plate come from a point 1 cm. from the bottom of the plate. The quality of tracks 2 and 3 is not high and they are so short that no measurement of their energies has been attempted. Tracks 4 and 5 are of better quality and are sufficiently long to permit of fairly accurate momentum measurements. The results of these measurements are surprising. They show that these particles are of opposite sign and have almost the same momenta (i.e. 10^9 ev/c.). These particles make angles of 35° and 46° respectively with the direction of the incident particle. Clearly, they are not electrons since two electrons of such high momenta are unlikely to emerge singly after traversing such a large thickness of lead. Again, the particles are unlikely to be heavier than protons. Thus, for each of the particles the ionization is typical of a fast particle and therefore the rest energies must be equal to or less than the momenta. Since the momenta are approximately 10^9 ev/c. the rest masses must be approximately equal to or less than the mass of the proton. If the particles are protons one is a negative proton. They may, however, be particles of intermediate mass. The large angle of emission does not help in deciding what kind of heavy particle, however, for with a suitable model a wide range of particles can be emitted at large angles to the direction of the incident particle. This shower raises many interesting questions, for example, what is the mechanism of the transfer of such a large amount of momentum, why are there so few low-energy particles emitted, and why are there so few mesons emitted? The answers to these questions await a fuller understanding of the interaction of fast nucleons.

§ 5. CONCLUSION

The following are the main results which follow from the analysis of the small sample of penetrating showers described in this paper. (i) A large fraction of the ionizing particles consists of particles which are different from ordinary mesons. Thus, out of twenty particles, with momenta greater than 10^9 ev/c., four are anomalously scattered in the 3 cm. lead plate and three produce no visible ionizing particle. (ii) A number of the positive particles are likely to be protons. This conclusion is suggested by the positive bias of the spectrum and by the presence of identifiable protons in the showers (for example, in shower 2 of Rochester and Butler 1947). (iii) The remaining penetrating particles may be either π -mesons (Lattes, Occhialini and Powell 1947) or heavier mesons (Rochester and Butler 1947). (iv) Some of the showers are very complex. There are complex showers which consist mainly of electrons and penetrating particles, and others which consist of large collimated groups of positively charged penetrating particles.

ACKNOWLEDGMENTS

It gives us great pleasure to express our sincere thanks to Professor P. M. S. Blackett, F.R.S. for the excellent facilities which he has given us and for the keen interest he has taken in this investigation. We are indebted also to Dr. S. M. Mitra and Mr. W. G. Rosser who operated the cloud chamber during part of the experiment and took several of the photographs described in this paper.

REFERENCES

- BLACKETT, P. M. S., 1937, *Proc. Roy. Soc. A*, **159**, 1.
- BRIDGE, H., HAZEN, W. E., and ROSSI, B., 1948, *Phys. Rev.*, **73**, 179.
- BROADBENT, D., and JÁNOSSY, L., 1947, *Proc. Roy. Soc. A*, **190**, 497.
- CHAKRABARTY, S. K., 1942, *Proc. Ind. Acad. Sci.*, **XV**, 472.
- CODE, F. L., 1941, *Phys. Rev.*, **59**, 229.
- DAUDIN, J., 1944, *Ann. Phys., Paris*, 11^e Sér., **19**, 110.
- FRETTER, W. B., 1948, *Phys. Rev.*, **73**, 41.
- HAZEN, W. E., 1944, *Phys. Rev.*, **65**, 67.
- JÁNOSSY, L., and ROCHESTER, G. D., 1943, *Proc. Roy. Soc. A*, **182**, 180.
- LATTES, C. M. G., OCCHIALINI, G. P. S., and POWELL, C. F., 1947, *Nature, Lond.*, **160**, 453, 486.
- LEPRINCE-RINGUET, L., and L'HERITIER, M., 1946, *J. Phys. Radium*, (Sér. 8), **7**, 66, 69.
- POWELL, W. M., 1946, *Phys. Rev.*, **69**, 385.
- ROCHESTER, G. D., 1946, *Proc. Roy. Soc. A*, **187**, 464.
- ROCHESTER, G. D., and BUTLER, C. C., 1947, *Nature, Lond.*, **160**, 855; 1948, *Proc. Phys. Soc.*, **61**, 307.
- ROCHESTER, G. D., BUTLER, C. C., and RUNCORN, S. K., 1947, *Nature, Lond.*, **159**, 227.
- SHUTT, R. P., 1946, *Phys. Rev.*, **69**, 261.
- WILSON, J. G., 1940, *Proc. Roy. Soc. A*, **174**, 73; 1946, *Nature, Lond.*, **158**, 414.

A few Remarks about Spectroscopy at Low Frequencies

By C. J. GORTER

Leiden

*MS. received 1 August 1948; read at Oxford Conference on
Microwave Spectroscopy 23 July 1948*

B EING the man who has just failed to make the important discoveries in radio- and microwave spectroscopy and having little to report about recent researches (Dr. Bloembergen will speak for himself) I shall say a few words about the reasons I had about twelve years ago to expect that it was possible to start spectroscopy at low frequencies.

It was clear that electric dipole lines would be broadened enormously by interaction and, moreover, often would not be allowed by the selection rules. So I came to consider the magnetic dipole transitions between the Zeeman levels of paramagnetic ions and between the nuclear Zeeman levels, and arrived at the following result.

If the absorption in the centre of the absorption band is described by the imaginary part χ'' of the magnetic volume susceptibility, we have $\chi'' \simeq \chi_0 \nu / \Delta \nu$, where $\Delta \nu$ denotes the width of the band and χ_0 the susceptibility connected with it. Often χ_0 is the static susceptibility, $\chi_0 = NM^2 / 3kT$, where N is the number of magnetic moments M per unit of volume. In relaxation bands $\Delta \nu \simeq \nu$, but in

narrow resonance lines we may win a factor 10^5 or more. For paramagnetic substances at room temperature we have $\chi_0 \simeq 10^{-5}$ and if $\Delta\nu \simeq \nu$ we get $\chi'' \simeq 10^{-5}$. For nuclei χ_0 is at least 10^5 smaller but we may win a large factor $\nu/\Delta\nu$ back for resonance lines.

The optical absorption coefficient in $I = I_0 \exp(-\alpha x)$, where I is the intensity in a plane wave, is

$$\alpha = 8\pi^2 \nu \chi'' / c, \quad \dots\dots(1)$$

which for $\chi'' = 10^{-5}$ and $\nu = 10^7$ becomes $\alpha \simeq 3 \times 10^{-7} \text{ cm}^{-1}$. Thus an absorption path of many kilometres would be needed for observation. In the normal optical region we are a factor 10^8 better off because of the larger ν while the extra advantage of the larger electric dipole probabilities is often outweighed by smaller densities.

At normal radio frequencies it is not difficult to measure $\chi'' = 10^{-5}$ either by bridge methods or by superheterodyne methods, while caloric measurements with an amplitude H_0 of 10 oersted permit observation of $\chi'' = 10^{-7}$ to 10^{-9} dependent on the temperature.

After the war it was found that with microwaves one could also easily reach $\chi'' = 10^{-5}$ while Purcell's refined bridges allow us to measure $\chi'' = 10^{-7}$ to 10^{-8} . The best accuracy at room temperature, however, was obtained by Zavoisky who apparently reached $\chi'' = 10^{-9}$ but who did not publish his method of observation in detail.

Finally I may quote from my old notes that in the case of a single spin the number n of transitions per second is

$$n = 4\pi M^2 H_0^2 / 3 \hbar^2 \Delta\nu. \quad \dots\dots(2)$$

Taking $\Delta\nu = 10^4$ and $H_0 = 10$ oersted we obtain $n = 2 \times 10^5 \text{ sec}^{-1}$ for hydrogen nuclei. This led to the suggestion of experiments with molecular beams.

Paramagnetic Resonance in Salts of the Iron Group— A Preliminary Survey: I. Theoretical Discussion*

By D. M. S. BAGGULEY, B. BLEANEY, J. H. E. GRIFFITHS,
R. P. PENROSE AND B. I. PLUMPTON

Clarendon Laboratory, Oxford

MS. received 27 August 1948

ABSTRACT. By means of paramagnetic resonance the energy levels of paramagnetic salts can be investigated in much greater detail than by any previous method. Using centimetre wavelength radiation it is possible to determine small splittings of the levels and their effective " g -value", and to isolate absorption lines due to different ions in unit cell. The interpretation of such effects in salts of the iron group by crystalline electric field theory is outlined.

The contributions from spin-spin and spin-lattice interaction to the widths of the lines are discussed. In the ions $3d_1\text{Ti}^{+++}$, $3d_2\text{V}^{+++}$, $3d_6\text{Fe}^{++}$ and $3d_7\text{Co}^{++}$, the spin-lattice relaxation time may be extremely short, and the lines correspondingly broad at room temperature; by working at sufficiently low temperatures, however, it should be possible to make the effects of spin-lattice interaction negligible in comparison with those of spin-spin interaction.

* This work was discussed at the Physical Society Conference at Oxford on 23rd–24th July 1948

§1. INTRODUCTION

THE use of a radio-frequency magnetic field to induce transitions between the different energy levels corresponding to the orientations of a nuclear spin in a magnetic field was first suggested by Gorter (1936), who attempted without success to find the corresponding absorption line. Using the molecular beam technique, Rabi and his co-workers (1939) applied the method to the determination of nuclear gyromagnetic ratios with outstanding results. Direct observation of resonance absorption and dispersion in liquids and solids was accomplished independently by Purcell, Torrey and Pound (1946) and Bloch, Hansen and Packard (1946). High precision measurements of nuclear g -values and much information on line breadths and thermal relaxation times have already resulted from this work.

Application of the same principles to electron spins instead of nuclear spins produces the analogous phenomenon of paramagnetic resonance. The energy absorbed by a paramagnetic salt from an oscillating magnetic field of a given frequency is measured as a function of a variable steady field applied at right angles to the alternating field. An absorption line is observed whenever the separation of the energy levels on the paramagnetic ion is equal to the quantum of energy corresponding to the radio frequency. By making measurements over a range of frequencies the transitions can be plotted as a function of the applied field. In this way the behaviour of the energy levels can be studied directly, and much more detailed information can be obtained than by any previous method, such as susceptibility and specific heat measurements, paramagnetic rotation and relaxation.

The purpose of this and the following paper is to present a preliminary survey of the method of paramagnetic resonance, with particular reference to salts of the first transition group. Measurements have been made on over a hundred salts, and in the first approximation, the results show that the determining factor is the number of electrons in the ion. These results, together with some discussion of each ion, are presented in the second paper. The first paper gives an outline of the effects of the crystalline electric field on the ions, together with a discussion of the question of line width, especially in connection with the spin-lattice relaxation time.

§2. THE ABSORPTION SPECTRUM

The fact that most paramagnetic salts of the iron group have susceptibilities conforming closely to those to be expected from electron spin only has been explained by the electric crystalline field theory developed by Bethe (1929), Van Vleck (1935) and others. Not only do the orbital and spin momentum vectors $\mathbf{L}$ and $\mathbf{S}$ fail to form a resultant $\mathbf{J}$, but the degeneracy of the $2L+1$ orientations of the orbital momentum is partly or wholly lifted according to the symmetry of the electric field. The separations between these orbital levels may amount to some 10^4 cm^{-1} , and the orbital momentum is therefore "quenched" at ordinary temperatures. The electron spin has no direct interaction with the electric field of the lattice, however, and therefore remains substantially "free", and the salts obey Curie's law (or a Weiss law) quite well at room temperature. In a magnetic field the degeneracy corresponding to the $2S+1$ orientations of the spin is lifted, and resonant absorption lines can be observed corresponding to transitions between these different levels. They may be detected by the damping of a tuned circuit containing the paramagnetic salt, as in the analogous nuclear resonance phenomenon. Owing to the large magnetic moment associated with the electron, the

energy levels will be correspondingly broader than in the nuclear case and are not sufficiently separated until fields of the order of kilogauss are applied. The absorption lines will lie then at centimetre wavelengths.

If the spin of the paramagnetic ion is perfectly free its $2S+1$ levels will have, in a magnetic field H , energies $-(g\beta H)M_S$, where M_S has the values $S, S-1, \dots -S$. Here g is the Lande " g -factor" and β is the Bohr magneton. When an alternating magnetic field is applied at right angles to H , transitions in M_S of the type $\Delta M_S = \pm 1$ are allowed. Equating the energy separation of adjacent levels to a quantum of energy ($hc\bar{\nu}$) one finds that an absorption line should occur at a wavelength of

$$\lambda = \frac{1}{\nu} = \frac{hc}{g\beta H} = \frac{4\pi mc^2}{e(gH)}. \quad \dots\dots(1)$$

If $g=2$, and H is measured in kilogauss, this may be written $H\lambda = 10.7$. Zavoisky (1945, 1946) and Cummerow, Halliday and Moore (1947) have shown that several salts of the iron group exhibit absorption maxima at centimetre wavelengths in magnetic fields close to that predicted by this equation.

In general the paramagnetic ion is not perfectly free, owing to a residual spin-orbit coupling, which produces the following effects: (a) the effective g -value is no longer exactly two; (b) when $S > \frac{1}{2}$, the levels may be split in zero magnetic field, with separations of the order of 1 cm^{-1} ; (c) the crystal is magnetically anisotropic, and the behaviour of the spin levels in a magnetic field depends on the orientation of the latter relative to the axes of the crystalline field; (d) when the unit cell of the crystal contains two or more paramagnetic ions, each subject to a differently oriented though otherwise identical crystalline field, it follows from (c) that the paramagnetic resonance spectrum of each ion will be different for an arbitrary orientation of the crystal.

When all these effects are present, the absorption spectrum of the salt will be very complicated and the interpretation difficult. In some cases sufficient data are available from x-ray measurements to deduce the nature of the crystalline field, and hence to calculate the behaviour of the energy levels in a magnetic field. A detailed comparison of experiment with theory is then possible, as, for example, in the case of chrome alum (Bagguley and Griffiths 1947, Weiss, Whitmer, Torrey and Hsiang 1947). These measurements show how strongly dependent the paramagnetic resonance spectrum is on the orientation of the crystal, even when the anisotropy of the susceptibility is very small. In cases where the exact nature of the crystalline field is not known, it is still possible to determine an initial splitting by plotting a transition back to zero magnetic field, as in ammonium chrome alum at low temperatures (Bleaney and Penrose 1948). It is not the purpose of this paper to consider particular salts in detail, however, but to assess the general advantages and limitations of the method of paramagnetic resonance in connection with the results reported in the following paper.

§ 3. THE BREADTH OF PARAMAGNETIC RESONANCE LINES

One of the chief limitations of the method of paramagnetic resonance lies in the lack of resolution of the fine structure of the spectrum owing to the breadth of the lines. The line broadening is associated with the interactions of the electron spins with each other and with the lattice; these interactions have been studied at lower frequencies by means of the phenomenon of paramagnetic relaxation by Gorte

and others (for a general review see Gorter 1948). The two interactions, spin-spin and spin-lattice, are somewhat different in nature and it is convenient to consider them separately.

Spin-spin interaction

In addition to the indirect effects of the crystalline field, each spin is subject to an interaction with the neighbouring spins, as a result of which the energy levels are spread out into a band. The continual exchange of energy between the various spins gives a lifetime for each spin in a given state of about 10^{-9} seconds, and at lower frequencies the phenomenon has been studied from the point of view of the non-resonant absorption connected with this relaxation time. Since paramagnetic resonance lines are generally observed at constant frequency in a variable magnetic field, it is more convenient to express their widths in terms of field, and the effect of spin-spin interaction may be regarded in the following way. Each ion is subjected not only to the external magnetic field, but also to the field of its neighbouring dipoles. In a weak field these are (classically speaking) randomly oriented, and there results a spread in the field acting on the individual dipoles, of the order μ/r^3 , where r is the mean distance between neighbouring dipoles. This "internal field" will give rise to a line width strongly dependent on the degree of "magnetic dilution"; even in such dilute salts as the alums the line width will amount to a few hundred gauss. One may expect a considerable change in line breadth with the orientation of the crystal in the applied magnetic field, as (to a first approximation) only the component of the local field in the direction of the applied field matters; the contribution from each neighbouring dipole to this will have an angular dependence of the form $(1 - 3 \cos^2 \theta)$ where θ is the angle between the line joining the dipoles and the applied field. In general, however, the width will be independent of the applied field (so long as saturation effects are negligible) and of the temperature.

The theory of the dipole-dipole interaction has been formulated by Van Vleck (1937) who has applied it to the calculation of line width in a forthcoming paper (Van Vleck 1948). An approximate method for the analogous nuclear case is outlined by Bloembergen, Purcell and Pound (1948). The calculation for paramagnetic salts is considerably more complicated, however, owing to the anisotropy and splitting of the energy levels caused by the crystalline electric field.

In addition to pure dipole-dipole coupling, it has generally been assumed that some kind of exchange interaction occurs between the ions. The theory of exchange forces in paramagnetic salts is in its very early stages, but it appears necessary to distinguish between an isotropic exchange and an anisotropic exchange force. The former is equivalent to a simple coupling between the spins proportional to the cosine of the angle between them, as in the Heisenberg theory of ferromagnetism. Gorter and Van Vleck (1947) point out that such exchange forces should cause a narrowing in the centre of the line and a broadening in the wings. Pryce (1948) has shown that in copper sulphate these forces, acting between the two different ions in unit cell can prevent the resolution of separate absorption lines due to the two ions, unless high magnetic fields are used. The mechanism of the anisotropic exchange is uncertain, though it has been applied by Opechowski (1948) to the explanation of the abnormally large specific heat of copper potassium sulphate below 1°K . discovered by de Klerk (1946).

Spin-lattice interaction

From the theory of paramagnetic relaxation it follows that when the magnetic field acting on a paramagnetic salt is changed, the magnetization M does not instantaneously take up the new value appropriate to the field, but approaches it exponentially with a time constant $\rho_L/2\pi$. The change of magnetization requires that the energy of the spin system change, which can only happen by means of an exchange of energy with the lattice. This exchange is, of course, continuous, so that the life of each paramagnetic ion in a given energy state is finite. This is true even when the external magnetic field is constant, since then the spins are only in a state of dynamic equilibrium which makes the total magnetic moment of the whole spin system constant when averaged over a time long compared to the relaxation time. When a radio-frequency field, whose period is comparable with or shorter than the relaxation time, is used in conjunction with a large constant field, the finite life time becomes manifest, though in different ways according to whether the alternating field is parallel or perpendicular to the constant field.

In the case of parallel fields, the addition of a small alternating field causes a first order fluctuation in the magnitude of the constant field, and hence, at low frequencies, a corresponding first order fluctuation of the magnetization. As the frequency increases, the spins are no longer able to exchange energy with the lattice sufficiently rapidly and the alternating component of the magnetization decreases. This decrease in the susceptibility is accompanied by a non-resonant absorption which is a maximum at a frequency equal to $(1/\rho_L)$. In the case of perpendicular fields, the vector resultant of a small alternating field and the large constant field contains only a second order fluctuation in magnitude. Consequently no exchange of energy between spin and lattice is required and there is no corresponding dispersion of the non-resonant type. The finite life-time of an ion in a given energy state due to spin-lattice interaction will, however, contribute to the width of the paramagnetic absorption line in addition to the width due to spin-spin interaction. When the spin-lattice relaxation time is long compared with the spin-spin relaxation time the additional width will be negligible, but when the former becomes comparable with the latter there will be a noticeable increase in line breadth. This point will be discussed in more detail later.

The effects of spin-lattice and spin-spin interaction may be summed up as follows:—

- Case 1. No constant magnetic field: non-resonant absorption at frequencies of the order $1/\rho_S$ (c/s.).
- Case 2. Constant magnetic field parallel to alternating field: non-resonant absorption at frequencies of the order $1/\rho_L$ and $1/\rho_S$ (c/s.).
- Case 3. Constant magnetic field perpendicular to alternating field: resonant absorption at frequencies corresponding to separation of the energy levels. Line breadth of the order $(1/\rho_S + 1/\rho_L)$ (c/s.).

§ 4. DISCUSSION

In certain cases it may happen that no paramagnetic resonance lines are observed, even at wavelengths and magnetic fields corresponding to the relation given in equation (1). There are two important reasons which must be considered in connection with this failure: (i) the spin levels may be split so that no two levels between which transitions are allowed have the right separation ($hc\bar{\nu}$) for the frequency of the measurement at a practicable value of the magnetic field; (ii) the

width of the absorption lines may be so great that no detectable change in the absorption occurs in the fields used. It is convenient to consider these two cases separately.

(i) By an important theorem of Kramers (1930), whenever an ion contains an odd number of electrons, the degeneracy of the spin levels can never be lifted entirely by an electric field. At least a two-fold degeneracy must remain, which can be resolved by a magnetic field. Paramagnetic resonance would therefore be expected in the ions with an even spin multiplicity, but in the remainder the separation in zero magnetic field may be too great to allow them to converge sufficiently in the fields at our disposal. The phenomenon should, however, appear at frequencies comparable with this initial splitting.

In this connection it is important to note that this splitting depends on λ^2/Δ where λ is the coefficient of the spin-orbit coupling $\lambda(\mathbf{L} \cdot \mathbf{S})$ and Δ is the splitting of the lowest orbital levels by the crystalline field (Van Vleck 1932).

(ii) The estimated line width H_1 due to spin-spin interaction amounts to some hundreds of gauss in hydrated paramagnetic salts. In the unhydrated salts the magnetic dipole field will be somewhat larger, but the effects of exchange forces are appreciable (e.g. metamagnetic curie points of anhydrous chlorides) and the paramagnetic resonance lines may be narrowed in consequence.* Thus the width due to spin-spin interaction should never be great enough to account for failure to detect resonance if magnetic fields up to 10 000 gauss are used.

The line width due to spin-lattice interaction is difficult to calculate. There is no direct coupling between the spins and the lattice, and two indirect couplings must be discussed. The first of these, originally pointed out by Waller (1932), acts through the magnetic dipole interaction of the spins, which varies periodically owing to the spatial vibrations of the ions under the action of the lattice waves. Through this coupling the spins can make transitions either by exchanging a quantum with a lattice vibration of the appropriate frequency, or by "scattering" a lattice quantum with change of energy ("Raman process"). Though the latter is a second order process, all lattice quanta can take part, whereas in the former only a limited number are available. In general the former process is the more important at helium temperatures, but at liquid oxygen temperatures the "Raman" processes should predominate.

The coupling mechanism considered by Waller leads to relaxation times longer than those observed experimentally, which in most salts are of the order of 10^{-8} second at 90°K . Further, in a number of cases, the failure to detect relaxation makes it necessary to suppose that the relaxation time is shorter than 10^{-8} second. An alternative coupling mechanism was proposed by Kronig (1939). Through the action of the lattice waves, fluctuating components of the crystalline electric field are set up which act on the orbital motion of the electrons, and hence cause transitions in the spin levels via the spin-orbit coupling. This mechanism is particularly potent when the orbital motion is only partly "quenched", that is, the splitting of the lowest orbital levels by the crystalline field is small. Using a

* An interesting example is anhydrous chromic chloride, CrCl_3 , on which measurements have been made at 3 cm. wavelength. The value of H_1 is about 1500 gauss, but even in the powder the half-width at half-intensity of the resonance line is about 50 gauss at room temperature! Since the crystals are probably anisotropic, somewhat narrower lines may be expected in a single crystal. At 20°K . the width is somewhat greater, but on cooling below the Curie point (16°K .) the line becomes very broad (about 2000 gauss) and the absorption is appreciable even in zero magnetic field.

simplified model, Kronig obtained the following estimates for the relaxation time in the case of $S = \frac{1}{2}$, the interchange of energy with the lattice vibrations being again either simple exchange of quanta or by the "Raman" process.

$$\text{Simple exchange: } \rho_L \simeq \frac{10^4 \Delta^4}{\lambda^2 H^4 T} (\text{sec.}) \quad \text{"Raman" scattering: } \rho_L \simeq \frac{10^4 \Delta^6}{\lambda^2 H^2 T^7} (\text{sec.})$$

In these formulae the temperature T is assumed small compared with the characteristic Debye temperature of the lattice, and Δ , λ are in cm^{-1} .

Since the effect of spin-lattice relaxation in broadening paramagnetic resonance lines is superimposed on that of spin-spin relaxation, for which the characteristic time is about 10^{-9} second, it follows that the effect will be negligible unless ρ_L is of the order of 10^{-9} second or shorter. To broaden the lines so much as to make them unobservable would require ρ_L to be as small as 10^{-11} second. No experimental data are available from which relaxation times of this order can be deduced, while the theory of Kronig is confined to the case of $S = \frac{1}{2}$. Van Vleck (1940) has made detailed calculations for chrome alum and caesium titanium alum, and for the "Raman" processes in the latter he obtains ($T \ll$ Debye lattice characteristic temperature) $\rho_L \simeq 10^{-15} \Delta^6 / T^9 (\text{sec.})$, independent of H so long as it is greater than the internal field. Thus if Δ were as small as 100 cm^{-1} , ρ_L would be 10^{-12} second even at 10°K . This case is probably exceptional, but from these calculations one may infer that the relaxation time will vary with a high power of the crystalline splitting of the lowest orbital levels, and with a high inverse power of the temperature. From this last it follows that by lowering the temperature it should be possible to make the spin-lattice relaxation time long enough to permit the observation of paramagnetic lines, and ultimately to reduce their width to that due to spin-spin relaxation alone.

To sum up, one may say that from both (i) and (ii) it appears that the salts in which paramagnetic resonance may not be observed are those in which the spin-orbit coupling is large, and more particularly those in which the splitting of the lowest orbital levels by the crystalline field is small. In general this means those paramagnetic ions in which an electric field of cubic symmetry leaves a degenerate orbital level at the bottom, the degeneracy being lifted only by a field of less symmetry. It should be possible to observe paramagnetic resonance even in these salts, however, by working at sufficiently high frequencies, when there are initial splittings, and at sufficiently low temperatures, when the spin-lattice relaxation time is abnormally short.

§ 5. SALTS OF THE IRON GROUP

The potential of the crystalline field is generally assumed to be of the form $V = Ax^2 + By^2 - (A+B)z^2 + D(x^4 + y^4 + z^4)$ though in some cases the axes of the second order (rhombic) field may be rotated with respect to those of the fourth order (cubic) field. Bethe (1929) has used group theory to show how the orbital levels split up in these fields, and an important point concerns which way up the multiplet appears, this depending for each ion on the sign of the coefficient D . Gorter (1932) has pointed out that if the ion is surrounded by six oxygen ions (negatively charged) arranged in an octahedron, the sign of D will be positive, whereas it will be negative if the arrangement is four in a tetrahedron or eight in a cube. In the following discussion the octahedral arrangement will be assumed, and the splitting of the orbital levels given accordingly.

The free ions of the first transition group are all in D-, F-, or S-states. In the latter case ($3d^6\text{Fe}^{+++}$, Mn^{++}) there is no orbital momentum to be quenched, and the crystalline field has only a very small effect on the spin levels. In a cubic field a D-state is decomposed into a doublet and a triplet, while an F-state resolves into a singlet and two triplets. The doublet state is non-magnetic, and when this is lowest, the splittings which are involved in the perturbation of the spin levels are those between the lowest level of the doublet and the triplet levels. Thus in this case, and in an F-state when the singlet lies lowest, the orbital levels are already sufficiently split up in a cubic field, the separation being of the order of 10^4 cm^{-1} , for the perturbation of the spin levels to be small. This applies to the ions $3d^3\text{Cr}^{+++}$, $3d^4\text{Cr}^{++}$, $3d^8\text{Ni}^{++}$ and $3d^9\text{Cu}^{++}$, in all of which paramagnetic resonance is observed at room temperature (see § 3 of the following paper) showing that the spin-lattice relaxation time is longer than 10^{-9} second.

The four remaining ions ($3d^1\text{Ti}^{+++}$, $3d^2\text{V}^{+++}$, $3d^6\text{Fe}^{++}$, $3d^7\text{Co}^{++}$) have a triplet level lowest, which will be split up in a field of low symmetry, but usually with a smaller separation than the ions considered previously. One may therefore expect a short spin-lattice relaxation time (Teunissen and Gorter 1939); in fact it has been found necessary to go to helium temperatures with Ti^{+++} , to hydrogen temperatures with Fe^{++} , and to oxygen temperatures with Co^{++} , to obtain resonance lines (no V^{+++} salt has been investigated).

In considering the question of the initial splitting of the spin levels it is not possible to draw such broad conclusions as in the last paragraph, except as regards the ions with Kramers' degeneracy. In all of these paramagnetic resonance has been detected at 10 cm. wavelength, using low temperatures where necessary. Of the other ions, one may mention Ni^{++} as a typical case, in most of whose salts no resonance has been detected at 3 cm. wavelength; this points to a rather large splitting ($>1\text{ cm}^{-1}$) as expected from previous work by other methods. Detailed consideration of the initial splitting is postponed, however, to the discussion of the individual ions in the following paper.

§ 6. CONCLUSION

It thus appears that all paramagnetic salts of the iron group should be open to investigation by the method of paramagnetic resonance, and one may note, in parenthesis, that the general considerations of this paper pertain also to salts of the rare earth group. The main difference lies in the fact that in the latter the crystalline field is not strong enough to destroy the coupling of the orbital and spin momenta to form a resultant $\mathbf{J}$, but does partially lift the degeneracy of its $2J+1$ orientations. The splittings are much larger than those of the spin multiplets in the iron group, and lie more in the far infra-red than at centimetre wavelengths. Modulation of the crystalline electric field by the Debye waves will now affect the paramagnetic ion directly and one may expect that the lattice relaxation times will be short, except at low temperatures. Few data are available on this point except in the case of gadolinium, to which, being in an ^8S state, these considerations do not apply.

In conclusion, one may say that paramagnetic resonance offers a method of detailed investigation of paramagnetic ions which possesses the following advantages over previous methods: (i) no corrections for diamagnetism or diamagnetic impurities; (ii) direct measurement of effective g values; (iii) direct measurement of small splittings of the order of 1 cm^{-1} ; (iv) direct observation of the different

paramagnetic ions in unit cell; (v) observation of very short spin-lattice relaxation times of the order of 10^{-10} second.

In addition it is to be expected that the interpretation of the spectra will yield valuable information concerning the crystalline electric field, and the shapes and widths of the lines will form a test of the theories of spin-spin interaction.

Full utilization of the method of paramagnetic resonance for investigating the behaviour of the spin energy levels under the combined influence of the crystalline electric field and an external magnetic field requires measurements on single crystals at various orientations, wavelengths and temperatures. Such measurements are being made on a number of salts, principally copper sulphate, the alums and the Tutton salts, and will be reported fully elsewhere. The experimental work presented in the following paper has been concerned mainly with powders, and measurements have been made of salts containing paramagnetic ions belonging to each of the nine spectroscopic states possible in the iron group (except $3d^2$), in order to make a preliminary survey to test the main conclusions of this paper.

REFERENCES

- BAGGULEY, D. M. S., and GRIFFITHS, J. H. E., 1947, *Nature, Lond.*, **160**, 532.
 BETHE, H. A., 1929, *Ann. Phys., Lpz.*, **3**, 133.
 BLEANEY, B., and PENROSE, R. P., 1948, *Proc. Phys. Soc.*, **60**, 395.
 BLOCH, F., HANSEN, W. W., and PACKARD, M., 1946, *Phys. Rev.*, **69**, 680.
 BLOEMBERGEN, N., PURCELL, E. M., and POUND, R. V., 1948, *Phys. Rev.*, **73**, 679.
 CUMMEROW, R. L., HALLIDAY, D., and MOORE, G. E., 1947, *Phys. Rev.*, **72**, 1233.
 GORTER, C. J., 1932, *Phys. Rev.*, **42**, 437; 1936, *Physica*, **3**, 995; 1948, *Paramagnetic Relaxation* (Amsterdam: Elsevier Publishing Co.)
 GORTER, C. J., and VAN VLECK, J. H., 1947, *Phys. Rev.*, **72**, 1128.
 DE KLERK, D., 1946, *Physica*, **12**, 513.
 KRAMERS, H. A., 1930, *Proc. Amsterdam Acad. Sci.*, **33**, 959.
 KRONIG, R. de L., 1939, *Physica*, **6**, 33.
 OPECHOWSKI, W., 1948, *Physica*, **14**, 237.
 PRYCE, M. H. L., 1948, *Nature, Lond.*, **162**, 539.
 PURCELL, E. M., TORREY, H. C., and POUND, R. V., 1946, *Phys. Rev.*, **69**, 37.
 RABI, I. I., ZACHARIAS, J. R., and KUSCH, P., 1939, *Phys. Rev.*, **53**, 318; **55**, 526.
 TEUNISSEN, P., and GORTER, C. J., 1939, *Physica*, **6**, 145.
 VAN VLECK, J. H., 1932, *Phys. Rev.*, **41**, 205; 1935, *The Theory of Electric and Magnetic Susceptibilities* (Oxford: Clarendon Press); 1937, *J. Chem. Phys.*, **5**, 320; 1940, *Phys. Rev.*, **57**, 426; 1948, *Phys. Rev.* (in the press).
 WALLER, I., 1932, *Z. Phys.*, **79**, 370.
 WEISS, P. R., WHITMER, C. A., TORREY, H. C., and JEN-SEN HSIANG, 1947, *Phys. Rev.*, **72**, 975.
 ZAVOISKY, E., 1945, *J. Phys. U.S.S.R.*, **9**, 211; 1946, *Ibid.*, **10**, 197.

Paramagnetic Resonance in Salts of the Iron Group— A Preliminary Survey: II. Experimental Results *

By D. M. S. BAGGULEY, B. BLEANEY, J. H. E. GRIFFITHS,
R. P. PENROSE AND B. I. PLUMPTON

Clarendon Laboratory, Oxford

MS. received 27 August 1948

ABSTRACT. Over a hundred salts of the iron group have been examined by the method of paramagnetic resonance at a wavelength of 3 cm., using magnetic fields up to 15 kilogauss at room temperature. The cases in which no absorption is observed have been further investigated at lower temperatures and shorter wavelengths. The results are considered for each paramagnetic ion in relation to the effects of the crystalline field theory. The ions Ti^{+++} , Fe^{++} and Co^{++} are found to have abnormally small spin-lattice relaxation times, while in Fe^{++} and Ni^{++} the energy levels are separated in zero magnetic field by as much as 1 cm^{-1} in most salts. A preliminary investigation of salts of the rare earth group indicates that spin-lattice relaxation times for most of these ions are also very short at temperatures above those of liquid hydrogen.

§ 1. INTRODUCTION

IN the preceding paper it has been shown that it should be possible to detect resonance in all paramagnetic salts of the iron group, but that the effect may not be manifest at all wavelengths and temperatures. It should be possible to ascribe any failure to two causes—too great an initial splitting of the energy levels, or too broad an absorption line due to a very short spin-lattice relaxation time. In the hydrated ions both these effects are closely related to the splitting of the orbital levels by the crystalline field, which is determined by the spectroscopic state of the ion, and by the arrangement of its surrounding molecules in the crystal. In spite of considerable variations in the latter, the crystalline electric field has in general predominantly cubic symmetry, and the success or failure in detecting paramagnetic resonance should be primarily associated with the spectroscopic state of the ion.

To test these conclusions, a preliminary survey of some hundred salts was made at room temperature, using a wavelength of three centimetres and magnetic fields up to 15 kilogauss. It was found that about forty salts failed to show any effect, and most of these have been further investigated at liquid oxygen and liquid hydrogen temperatures and at shorter wavelengths. In this paper the results of these measurements are presented and analysed, ion by ion, in terms of the general discussion of the preceding paper.

§ 2. THE EXPERIMENTAL METHOD

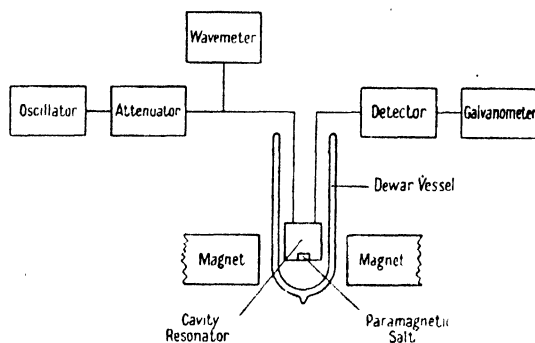
In this section a brief outline of the essentials of the experimental method will be attempted, without any detailed discussion of the technique.

The paramagnetic resonance line is detected in absorption by the damping of a cavity resonator. At the shorter wavelengths this consists generally of a section of circular waveguide excited in the H_1 mode, the crystal being placed at one end in a position of maximum magnetic field. At longer wavelengths either a rectangular

* This work was discussed at the Physical Society Conference at Oxford on 23rd–24th July 1948

waveguide resonator is used, excited again in the lowest H mode, or a quarter-wavelength section of coaxial line open-circuited at one end and close-circuited at the other. An external magnetic field is applied perpendicularly to the radio-frequency magnetic field.

The magnification factor Q of the empty cavity is normally several thousand, and Kittel and Luttinger (1948) have shown that in an average case this would be reduced to a few hundred at the centre of a paramagnetic absorption line, if the cavity is filled with salt. This arrangement is impracticable for measurements on single crystals and, moreover, most of the salts to be investigated show strong dielectric absorption at centimetre wavelengths. This dielectric absorption is minimized by placing the crystal at a node in the electric field, but owing to its finite thickness the crystal will project slightly into the electric field. The amplitude of the electric field increases sinusoidally with distance from the end, and it is easy to show that the optimum thickness of crystal (i.e. that producing maximum paramagnetic absorption in proportion to other losses in the resonator) is that whose dielectric losses reduce the Q to one-third of the value for the empty cavity. Under these conditions the reduction in the Q due to the paramagnetic resonance is relatively small, and is measured by the change in the power transmitted through the cavity from an oscillator to which it is tuned (see figure).



Block diagram of apparatus.

This power is measured by a silicon-tungsten crystal rectifier which, when operated at low power, gives a rectified current through a low resistance galvanometer closely proportional to the R.F. power. Then the paramagnetic absorption can be calculated from the formula

$$\text{Fractional increase in } 1/Q = (\delta_0/\delta_1)^4 - 1, \quad \dots\dots (1)$$

where δ_1 , δ_0 are the galvanometer deflections in the presence and absence of paramagnetic resonance absorption respectively.

The method of measurement may be summarized as follows. A crystal of the paramagnetic salt is placed in the cavity, which is tuned to an oscillator of the desired frequency. An external magnetic field is then applied, and slowly varied in magnitude until the resonance is observed. The absorption curve due to the salt is then delineated by plotting the quantity given in equation (1) as a function of the external field. At each setting of the field the tuning either of the cavity or the oscillator is adjusted to give maximum galvanometer deflection, to avoid errors due to drift of the oscillator in frequency, to detuning of the cavity by the anomalous dispersion in the paramagnetic salt, or to distortion caused by magnetic forces.

The experimental technique at low temperatures is fundamentally the same, though rather more complicated than at room temperature; this drawback is outweighed in many cases by a number of advantages. Apart from the necessity of using low temperatures when the spin-lattice relaxation time is too short at room temperature, the advantages are, in order of importance: (a) the absorptions are much stronger, corresponding to the increase in susceptibility according to Curie's law; (b) the damping of the resonator due to dielectric losses in the crystal is usually, though not invariably, negligible; (c) the Q of the empty cavity is somewhat higher.

Each of these effects enhances the absorption due to the paramagnetic resonance in comparison with other losses; this is especially valuable when it is difficult to grow large single crystals, and in studying the shapes of resonance lines in the wings.

§ 3 RESULTS

The great majority of the salts were measured first at room temperature at a wavelength of 3 cm., using magnetic fields up to 15 kilogauss. It is convenient to present the results of this work first, and then to discuss the success or failure to detect resonance in the light of further experiments (§ 4). Table 1 contains therefore a list of the salts investigated, grouped according to the spectroscopic state of the paramagnetic ion, with a dividing line to separate those salts in which no absorption due to paramagnetic resonance could be detected under the conditions: temperature 290° K., wavelength 3.2 cm., magnetic field, varied from zero to 15 kilogauss (see p. 554).

Before proceeding further it is convenient to eliminate from the list of failures those which may be ascribed to a reason trivial from the theoretical standpoint, namely, that the dielectric losses in the crystals masked the paramagnetic absorption. These losses were particularly great in the hygroscopic salts, owing to the enormous absorption in water at centimetre wavelengths. The latter disappears when the water is frozen, and, in general, the dielectric losses in the crystals themselves decrease markedly at low temperatures. Salts in which the failure to detect paramagnetic resonance is ascribed to this effect are marked with an asterisk, those salts in which the resonance appears at lower temperature owing to the decrease in dielectric losses are marked with a double asterisk. It should be noted also that susceptibility measurements indicate that the cobalt-ammine complexes are diamagnetic.

Of the salts in which the failure to observe resonance could not be ascribed to dielectric loss, a considerable number have been further investigated at shorter wavelengths and lower temperatures. The results are summarized in table 2, where the conditions under which resonance was detected are given at the foot of each column (see p. 555).

§ 4. REVIEW OF THE INDIVIDUAL IONS

A brief general review of the effects of the crystalline field on the orbital levels of the ions of the first transition group has been given in § 5 of the preceding paper. In the light of this review, the results just presented merit some further discussion, which is most conveniently assembled separately for each ion. As before, the sign of the coefficient D of the fourth order portion of the crystalline field potential will be taken to be positive, except where otherwise stated. The magnitude and sign of the coefficient λ of the spin-orbit coupling play an important rôle and are taken from the calculations of Laporte (1928).

Table 1. Salts investigated at 3 cm. wavelength, room temperature

Resonance observed					
$4F_{3,2}$	$5D_0$	$6S$	$3F_4$	$3D_{3,2}$	
VS_4	$CrCl_3 \cdot 6H_2O$	MnF_2	$NiCl_2$	CuF_2	
CrF_3	$CrCl_3 \cdot 4H_2O$	$MnCl_2 \cdot 4H_2O$	$NiBr_2$	$CuCl_2 \cdot 2H_2O$	
$CrCl_3$	$CrCl_3 \cdot 4H_2O$	$MnSO_4 \cdot 4H_2O$		$CuSO_4$	
$CrCl_3 \cdot 6H_2O$		$K_2Mn(SO_4)_2 \cdot 6H_2O$		$CuSO_4 \cdot 5H_2O$	
$Cr_2(SO_4)_3 \cdot 3H_2O$		$Mn(NH_4)_2(SO_4)_2 \cdot 6H_2O$		$Cu(NH_4)_2SO_4 \cdot H_2O$	
$Cr_2(SO_4)_3 \cdot 15H_2O$		$Mn(C_2H_3O_2)_2 \cdot 4H_2O$		$Cu(NH_4)_2Cl_2 \cdot H_2O$	
$KCr(SO_4)_2 \cdot 12H_2O$		$Mn_2(AsO_4)_2$		$CuCl_2 \cdot 2NH_4Cl \cdot 2H_2O$	
$NH_4Cr(SO_4)_2 \cdot 12H_2O$		$Mn_2(C_2H_3O_2)_2$		$CuK_2(SO_4)_2 \cdot 6H_2O$	
$RbCr(SO_4)_2 \cdot 12H_2O$		$Mn_2Fe(CN)_6 \cdot 7H_2O$		$CuCs_2(SO_4)_2 \cdot 6H_2O$	
$(NH_4CH_3)Cr(SO_4)_2 \cdot 12H_2O$		$Mn_2(PO_4)_2 \cdot 7H_2O$		$Cu(NH_4)_2(SO_4)_2 \cdot 6H_2O$	
$Cr(NO_3)_3 \cdot 9H_2O$		$Mn(CHO)_2 \cdot 2H_2O$			
$CrPO_4 \cdot 3H_2O$		$Mn_2(P_2O_7) \cdot 3H_2O$			
$Cr(C_2H_3O_2)_3 \cdot H_2O$		$(MnC_2O_4)_2 \cdot 5H_2O$			
$Cr(OH)_3$		$MnCO_3$			
$Cr_2(CrO_4)_3$		$Mn(H_2PO_4)_2 \cdot H_2O$			
$CsCr(SO_4)_2 \cdot 12H_2O$		$Mn(C_2H_3O_2)_3 \cdot 3H_2O$			
		$(FeF_3)_2 \cdot 9H_2O$			
		$FeCl_3$			
		$Fe_2(SO_4)_3 \cdot 9H_2O$			
		$FeNH_4(SO_4)_2 \cdot 12H_2O$			
		$FeOH(C_2H_3O_2)_2$			
		$Fe(C_2H_3O_2)_3$			
		$Fe_2(FeCN)_6$			
		$Fe(PO_4) \cdot 4H_2O$			
		$Fe_2[C_2H_3(OH)_2OPO_3]_3$			
		$Fe_2(C_2O_4)_3$			
		No resonance observed			
	$5D_4$	$4F_{9,2}$	$3F_4$	$3D_{5/2}$	
$3D_{3,2}$					
$CsTi(SO_4)_2 \cdot 12H_2O$	$FeCl_3$	$CoCl_2 \cdot 6H_2O$	NiF_2	$CuBr_2^{**}$	
	$FeCl_2 \cdot 4H_2O$	$CoSO_4 \cdot 7H_2O$	$NiCl_2$	$Cu_3(PO_4)_2 \cdot 3H_2O$	
	$FeSO_4 \cdot 7H_2O$	$Co(NO_3)_2 \cdot 6H_2O$	$NiCl_2 \cdot 6H_2O$		
	$Fe(NH_4)_2(SO_4)_2 \cdot 6H_2O$	$Co(NH_4)_2(SO_4)_2 \cdot 6H_2O$	$NiSO_4 \cdot 7H_2O$		
	$K_2SO_4FeSO_4 \cdot 6H_2O$	$NH_4ClCoCl_2$	$K_2Ni(SO_4)_2 \cdot 6H_2O$		
	$(FeSO_4)_2 \cdot 5H_2O$	$Co(H_2O)(NH_4)_6Cl_3$	$Ni(NO_3)_2 \cdot 6H_2O$		
	$Fe_2(PO_4)_3 \cdot 8H_2O$	$[Co(NH_3)_6Cl]Cl_3$	$Ni_3(PO_4)_2 \cdot 7H_2O$		
	$Fe_2O_4 \cdot 2H_2O$	$[Co(NH_3)_4(CO_3)]NO_3 \cdot 6H_2O$	$Ni(CHO)_2 \cdot 2H_2O$		
	$FeBr_2 \cdot 6H_2O$	Cs_2CoCl_6	$Ni(NH_4)_2(SO_4)_2 \cdot 6H_2O$		
	$Co_2 \cdot O_3$				

Headings of columns indicate spectroscopic state. For meaning of asterisks see text.

Table 2. Substances further investigated at lower temperatures and shorter wavelengths

T (° K.) λ (cm.)	Resonance observed				T (° K.) λ (cm.)
	Ti ⁺⁺⁺	Fe ⁺⁺⁺ Mn ⁺⁺	Fe ⁺⁺	Ni ⁺⁺	
6 10	CsTi(SO ₄) ₂ 12H ₂ O	K ₃ Fe(CN) ₆	FeK ₂ (SO ₄) ₂ ·6H ₂ O	NiCl ₂ ·6H ₂ O Ni(NH ₄) ₂ (SO ₄) ₂ ·6H ₂ O NiSiF ₆ ·6H ₂ O	90, 290 1·3, 0·8
			Fe(NH ₄) ₂ (SO ₄) ₂ ·6H ₂ O	CoCl ₂ ·6H ₂ O Co(NO ₃) ₂ ·6H ₂ O CoSO ₄ ·7H ₂ O CoK ₂ (SO ₄) ₂ ·6H ₂ O Co(NH ₄) ₂ (SO ₄) ₂ ·6H ₂ O Cs ₃ CoCl ₅	
90 3			90 3		
			No resonance observed		
T (° K.) λ (cm.)	Mn ⁺⁺		Co ⁺⁺	Ni ⁺⁺	T (° K.) λ (cm.)
	Mn(NH ₄)PO ₄ ·H ₂ O		Co(H ₂ O)(NH ₃) ₅ Cl ₃ NH ₄ CoCl ₃	NiF ₂ NiI ₂ NiSO ₄ ·7H ₂ O Ni(NO ₃) ₂ ·6H ₂ O NiK ₂ (SO ₄) ₂ ·6H ₂ O	
90 3			90 3	90 1·3	

T = temperature. λ = wavelength.

Headings of columns indicate spectroscopic state.

$3d^1$, ${}^2D_{3/2}Ti^{+++}$, $\lambda = 154\text{ cm}^{-1}$

The five-fold orbital level is split in a cubic electric field into a triplet and a doublet, the former being the lower when the sign of the coefficient D is positive. The degeneracy of the triplet is lifted by a rhombic field, but the exact splitting is not known. The spin is $\frac{1}{2}$, and no electric field can separate the two spin levels by Kramers' (1930) theorem. The absence of paramagnetic resonance in caesium titanium alum must therefore be due to a very short spin-lattice relaxation time; this is consistent with negative results of relaxation experiments (Gorter, Teunissen and Dijkstra 1938, de Haas and du Pré 1938). No resonance was detected at 20°K . and 3 cm. wavelength, but Bijl (unpublished) has since found in this laboratory that it appears at 6°K ., using a wavelength of 10 cm. Such a short relaxation time could be explained by the theory of Van Vleck (1930), if the orbital splitting were as low as 100 cm^{-1} , the dominant effect being the "Raman process". The effective g value is considerably less than two, as would be expected from the positive spin-orbit coupling and the small orbital splitting.

$3d^2$, ${}^3F_2V^{+++}$

In a cubic field the seven-fold orbital level splits into a singlet and two triplets, with a triplet lowest, which will be split in a field of lower symmetry (see Siegert 1936, 1937). The spin triplet is split by a trigonal field into a singlet ($M_s = 0$) and a doublet ($M_s = \pm 1$), with a separation of 5 cm^{-1} , according to susceptibility measurements on vanadium ammonium alum of van den Handel and Siegert (1937) (see also Van Vleck 1939a). Except for weak transitions between $+1$ and -1 levels, one would not expect paramagnetic resonance to be detectable except at low temperatures and millimetre wavelengths. No measurements have yet been made.

$3d^3$, ${}^4F_{3/2}Cr^{+++}$, $\lambda = 87\text{ cm}^{-1}$; V^{++} , $\lambda = 55\text{ cm}^{-1}$

In a cubic field the orbital singlet term lies lowest, some 10^4 cm^{-1} below the nearest triplet. The spin quadruplet remains completely degenerate in a cubic field, but is resolved into two doublets by any field of lower symmetry, the Kramers degeneracy always being present. Thus all Cr^{+++} and V^{++} salts should show paramagnetic resonance at room temperature at all wavelengths. The separation of the two spin doublets is generally rather small (a fraction of 1 cm^{-1}), and magnetically dilute Cr^{+++} salts such as the alums have been widely used for adiabatic demagnetization experiments below 1°K . Some measurements at room temperature by the method of paramagnetic resonance have already been reported by Bagguley and Griffiths (1947) and by Weiss, Whitmer, Torrey and Hsiang (1947), who find a separation of 0.12 cm^{-1} for the potassium alum and 0.15 cm^{-1} for the ammonium alum. Bleaney and Penrose (1948b) have found that as the salts are cooled down the splittings change in a marked fashion. Complete reports of measurements, both at room temperature and at low temperatures, on these and three other chrome alums will be presented shortly.

The abnormally narrow line obtained with $CrCl_3$ has already been commented on in a footnote of the preceding paper. The fact that only a single line is obtained is probably due to exchange forces strong enough to overcome the effects of the crystalline electric field.

Only one divalent vanadium salt has been examined: the sulphate. This gave a fairly narrow absorption line close to the field for $g=2$.

$3d^4$, $^5D_0Cr^{++}$, $\lambda = 57\text{ cm}^{-1}$

In a cubic field the orbital levels split, like those of copper into a lower "non-magnetic" doublet and an upper triplet. As regards the perturbation of the spins, only the rather large energy separations between the lower level of the doublet and the triplet levels play any part, since the orbital doublet is non-magnetic. Also, the spin-orbit coupling coefficient is rather small, and one may expect the initial splittings of the spin quintuplet to be small, and the spin-lattice relaxation time to be fairly long. These predictions are confirmed by the observation of paramagnetic resonance at room temperature in the salts investigated.

$3d^5$, $^6S_{5/2}Fe^{+++}$, Mn^{++}

These ions are in S states when free, and it follows that effects of the crystalline electric field are very small indeed. The spin sextuplet is resolved into a doublet and a quadruplet by a cubic field, and into three doublets by a trigonal field. As expected, resonance was detected in practically all the salts examined. A special case worthy of note is potassium ferricyanide, where the bonding is partly covalent and the much stronger electric fields thereby produced are able to break down the Russell-Saunders coupling (Howard 1935). The spins pair off as far as possible, making the ferrocyanide diamagnetic, while in the ferricyanide the susceptibility approaches that of a single spin. In an experiment at 20°K . and 3 cm. wavelength an exceedingly broad absorption was observed, increasing with field, but the maximum was not reached at 8000 gauss.

Single crystals have been investigated of two salts used in low temperature adiabatic demagnetization work: iron ammonium alum and manganese ammonium sulphate. Owing to the large spin the line breadth is so great that accurate measurements of splittings are impossible. In the alum the splitting appears to be not more than 0.10 cm^{-1} ; this conflicts with low temperature measurements of Benzie and Cooke (1948), which give 0.14 cm^{-1} , assuming a cubic field. The explanation may lie either in a trigonal field which decomposes the energy levels into three doublets of small separation, or in a magnetic specific heat larger than that calculated from dipole-dipole interaction, which would correspondingly reduce the specific heat attributed to the Stark splitting and hence the estimate of the splitting from low temperature measurements. In the manganese Tutton salt the experiments suggest that the energy levels are split into three doublets by the crystalline electric field, with separations of the order of 0.1 and 0.2 cm^{-1} .

$3d^6$, $^5D_4Fe^{++}$, $\lambda = -100\text{ cm}^{-1}$

In a cubic field the orbital levels split to a basic triplet and an upper doublet, being fully resolved by a rhombic field. The latter will also fully resolve the spin quintuplet, though fields of higher symmetry leave some degeneracy. There appear to be practically no data from which an estimate of the splittings can be made but one may expect them to be fairly large, and the spin-lattice relaxation time to be short. No resonance has so far been detected at 3 cm. using fields up to 8 kilogauss at temperatures down to 90°K ., though a weak absorption appears at 20°K . Measurements at 1.3 cm. wavelength have revealed strong resonance lines in ferrous ammonium and ferrous potassium sulphate at 20°K ., though nothing could be detected at 90°K . The potassium salt showed strong absorption in zero field, indicating that one of the splittings of the spin quintuplet must be about 0.8 cm^{-1} in this case.

$3d^7$, ${}^4F_{9/2}Co^{++}$, $\lambda = -180\text{ cm}^{-1}$

The orbital levels are split up in the same way as those of V^{+++} , but the spin levels can only be resolved into two doublets by an electric field. Their separation will be much greater than that in Cr^{+++} , owing to the larger value of λ and the smaller orbital splitting, and Van Vleck (1939 b) has suggested that the splitting may be as much as 10 cm^{-1} . It should be possible, however, to detect paramagnetic resonance due to transitions between the $\pm \frac{1}{2}$ levels, and the negative results at room temperature must be due to a short spin-lattice relaxation time, as would be expected from the small orbital splitting. Measurements at lower temperatures have confirmed this, broad lines being obtained at 90° K . With single crystals of the Tutton salts, fairly narrow lines are observed at 20° K . with effective g values as high as three, and very considerable anisotropy. The spectrum is more complicated than would be expected with an effective spin of $\frac{1}{2}$, and the theoretical analysis may be rather involved (Schlapp and Penney 1932). The increased width of the lines at 90° K . suggests that the spin-lattice relaxation time is about 10^{-9} seconds at this temperature.

The cobalt salts shown in table 1 (except Cs_3CoCl_5) are all pink-coloured, showing that the cobalt ion has six coordination bonds, corresponding to a positive value of the coefficient D of the cubic field. In the blue-coloured salts the coordination number is four, and the sign of D should be reversed. A singlet orbital level should then be lowest, as in Cr^{+++} in a cubic field. These blue salts should then show paramagnetic resonance at room temperature, since the spin-lattice relaxation time should not be unduly short, and g values close to two. Krishnan and Mookherji (1938) have found that Cs_3CoCl_5 , which Powell and Wells (1935) have shown to be four-coordinated, exhibits less anisotropy and its mean susceptibility is somewhat closer to the "spin-only" value than most cobalt salts. Investigation by means of paramagnetic resonance reveals no significant difference in behaviour however; it is necessary to cool the salt to 90° K . to obtain a narrow resonance curve, and the g value does not appear significantly smaller than in other cobalt salts.

$3d^8$, ${}^3F_4Ni^{++}$, $\lambda = -335\text{ cm}^{-1}$

In a cubic field the orbital levels of Ni^{++} are similar to those of Cr^{++} , but inverted to those of Co^{++} and V^{+++} ; the lowest level is therefore a singlet. The spin triplet is decomposed by a non-cubic field as in V^{+++} , but the splitting is smaller. Schlapp and Penney (1932) estimate its splitting to be 1.5 cm^{-1} , but in fitting their calculations to the measurements of Bartlett on the Tutton salts they ignore the presence of two ions in unit cell and assume that the axes of the crystalline field are the same as those of the magnetic susceptibility. From specific heat measurements at low temperatures Giauque and Stout (1941) obtain an overall splitting of 5.3 cm^{-1} for $NiSO_4 \cdot 7H_2O$, assuming three almost equally spaced levels. Broer, Dijkstra and Gorter (1943) point out that the measurements of Krishnan and Mookherji (1938) on anisotropy in this salt and $Ni(NH_4)_2(SO_4)_2 \cdot 6H_2O$ lead to splittings two to three times larger than they obtain from paramagnetic relaxation: the latter gives 1.5 cm^{-1} for both salts (assuming a doublet and a singlet level). There is thus considerable uncertainty in the value of the splitting in Ni^{++} salts.

Measurements at a wavelength of 1.3 cm . on nickel ammonium sulphate have revealed a paramagnetic resonance line in rather high fields (about 12 kG.) though nothing could be detected in nickel potassium sulphate. This suggests that the

splitting in the former salt is not much greater than 1 cm^{-1} , resonance being detected at a lower frequency (0.8 cm^{-1}) because one of the $M_s = \pm 1$ energy levels may converge in a magnetic field on the $M_s = 0$ level. No resonance was observed at 3 cm. at temperatures down to 20°K ., though it is possible that a line may appear if a sufficiently large magnetic field could be applied to move the energy levels still closer. Evidently the transition between the $+1$ and -1 levels, assuming them to be degenerate in zero magnetic field is very weak, since it is not detectable at 20°K .

The detection of paramagnetic resonance in two unhydrated nickel halide salts shows that either the crystalline field is too small to produce a splitting of 0.3 cm^{-1} , or that exchange forces arise powerful enough to overcome the effects of the crystalline field. (NiCl_2 has a Curie point at 50°K .). No resonance lines could be detected in the anhydrous fluoride or iodide at either 3 cm. or 1.3 cm. wavelength, even at 90°K .

A hydrated salt in which the splitting of the triplet is small is $\text{NiSiF}_6 \cdot 6\text{H}_2\text{O}$. From measurements of paramagnetic rotation of the plane of polarization of light by Becquerel and van den Handel (1939) down to liquid helium temperatures, a splitting of 0.3 cm^{-1} was deduced by Becquerel and Opechowski (1939). Since this crystal contains only one ion per unit cell, it is particularly simple for study by the paramagnetic resonance method. Measurements by one of us (R.P.P.) have shown that the splitting is 0.32 cm^{-1} at 195°K ., decreasing as the temperature is lowered to 0.17 cm^{-1} at 90°K . and 0.12 cm^{-1} at 20°K .; no further change occurred down to 14°K . Since the splitting is entirely due to the non-cubic part of the crystalline field, the change is probably due to anisotropic thermal contraction, which should vanish near the absolute zero. The g value is 2.28.

$3d^9$, $^2D_{5/2}\text{Cu}^{++}$, $\lambda = -852\text{ cm}^{-1}$

In a cubic field, the orbital levels split into a "non-magnetic" doublet and triplet, the former being the lower. Thus one may expect a fairly long spin-lattice relaxation time, and in view of the Kramers' degeneracy practically all hydrated copper salts should show paramagnetic resonance at room temperature at 3 cm. wavelength. The spin-orbit coupling coefficient is very large however, and the effective g values are appreciably greater than two, being accompanied by considerable magnetic anisotropy. Single crystals of copper sulphate and the copper Tutton salts have been investigated, and the results obtained with the latter are in good agreement with the theory of Polder (1942). In the case of $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$, where the effects of exchange forces are pronounced, it is necessary to use millimetre wavelengths (i.e. magnetic fields $>10\text{ kG}$.) to separate the absorption lines associated with the two ions in unit cell (Bagguley and Griffiths 1948, Pryce 1948).

§ 5. DISCUSSION: LINE BREADTH

Though the presence of magnetic anisotropy can be detected by measurements of susceptibility on single crystals, its effects are more marked in paramagnetic resonance than in any other phenomenon. In nickel ammonium sulphate for instance, a strong resonance line was observed at 1.3 cm. wavelength in one orientation with respect to the magnetic field, whereas only a very slight absorption could be detected on turning the crystal through a right angle. Thus it was desirable as well as convenient to make most of the measurements in this preliminary survey on powders, in which all orientations are present simultaneously. The resultant absorption is then rather broad, consisting of many lines superimposed, and little can be said about the shape or width of the individual lines. In a number of cases observations have been made on single crystals, which have shown that the line

shape and width depend very much on the orientation and on the crystal. Even in the cases where the width is ascribed to the effects of spin-spin interaction alone, no general expression for the shape can be given. Lines have been observed (in caesium chromium alum) whose shape conforms closely to a Gaussian distribution, as assumed by Broer (1943) in his theory of spin-spin relaxation, and also (in $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$) to the simple Lorentz pressure-broadening shape for spectral lines in gases; but these are exceptional cases, and a variety of intermediate shapes have been found. The complexity of the problem is illustrated by the copper Tutton salts, where the maximum and minimum widths for different orientations of the crystal (themselves different by a factor three) vary by a factor three between salts with different monovalent metallic ions.

In salts where spin-lattice relaxation is the dominant effect, it is reasonable to assume that the line shape (as a function of frequency) should be given by an expression of the form

$$F(\nu) = \left\{ \frac{\Delta\nu}{\Delta\nu^2 + (\nu - \nu_0)^2} + \frac{\Delta\nu}{\Delta\nu^2 + (\nu + \nu_0)^2} \right\},$$

where $\Delta\nu = 1/\rho_L$, since the assumption that the time between "collisions" with the lattice during which an ion can absorb energy from the radiation field has an exponential distribution of the type $f(t) = \exp(-2\pi t/\rho_L)$ is the same as in the theory of Van Vleck and Weisskopf (1945) and Fröhlich (1946) for pressure broadening. If paramagnetic resonance lines could be observed at constant magnetic field (i.e. constant ν_0) and the absorption curve delineated as a function of frequency, it would be a simple matter to deduce the relaxation time. As they are in fact observed at constant frequency and at variable field, it is difficult to obtain more than an estimate of the relaxation time from a single measurement since in general ρ_L is a function of the external magnetic field, and the line shape is obtained as a function of the field and not of the frequency. By making measurements at different frequencies it should be possible to interpret the curves, however, and obtain the spin-lattice relaxation time at different temperatures and fields.

In the preceding paper it was suggested that the ions of the rare earth group (except Gd^{+++}) might have very short spin-lattice relaxation times. Some preliminary experiments to test this were made with a few salts kindly lent to us by Mr. J. R. Marsh of the Inorganic Chemistry Laboratory, Oxford. At room temperature and 3 cm. wavelength only gadolinium (bromate and acetate), terbium (bromate) and europium (chloride) showed any absorption. A further investigation was therefore made at low temperatures, using a wavelength of 10.4 cm. It was found that all the ions examined (cerium, neodymium, samarium, europium, terbium, dysprosium, erbium and ytterbium), showed resonance lines at 20° K., but nothing could be detected at 90° K. except in europium. In general, then, the spin-lattice relaxation time is very short at high temperatures, as expected from the strong coupling between spin and orbit. The case of the europium salt, $\text{EuCl}_3 \cdot 6\text{H}_2\text{O}$, is very puzzling, as the absorption was considerably stronger at 20° K. than at 90° K., whereas all the ions should be in the lowest non-magnetic state ($J=0$) at this temperature.

When the spin-lattice relaxation time is long compared with the spin-spin relaxation time, its contribution to the widths of the lines is negligible. Nevertheless if sufficient power can be provided to develop the saturation phenomenon, as found in the nuclear magnetic resonance experiments (see, for example, Bloembergen, Purcell and Pound 1948) and in the inversion spectrum of ammonia

(Bleaney and Penrose 1948 a) it should be possible to measure long spin-lattice relaxation times, either by applying short pulses and observing the variation of the paramagnetic absorption with time on an oscilloscope or by measuring the diminution in the equilibrium value of the absorption as the radio-frequency power level is increased. At low temperatures, where the spin-lattice relaxation time increases rapidly, the microwave power required is quite small, of the order of milliwatts when $\rho_L > 10^{-3}$ second. It is difficult at this stage to estimate the merits of this type of measurement in comparison with the conventional methods of paramagnetic relaxation.

§ 6. CONCLUSION

The chief merits of the method of paramagnetic resonance for the investigation of the energy levels of paramagnetic ions have been summarized at the end of the preceding paper, where it was pointed out that measurements on single crystals are essential to make full use of the method. A few such measurements have been mentioned in this paper, but the bulk of the observations have been made on powders in order to obtain a fairly rapid survey of the iron group. The results indicate that the main features of the resonance phenomenon are consistent with the crystalline field theory in its broad outline. More detailed measurements will show how far the finer points of the theory are supported by experiment, and should throw light on the nature of the spin-spin interaction.

ACKNOWLEDGMENT

The authors' thanks are due to Professor the Lord Cherwell for extending to us the facilities of the Clarendon Laboratory, and to Professor F. Simon and his co-workers for the supply of liquid hydrogen.

REFERENCES

- BAGGULEY, D. M. S., and GRIFFITHS, J. H. E., 1947, *Nature, Lond.*, **160**, 532; 1948, *Ibid.*, **162**, 538.
 BECQUEREL, J., and VAN DEN HANDEL, J., 1939, *Physica*, **6**, 1034.
 BECQUEREL, J., and OPECHOWSKI, W., 1939, *Physica*, **6**, 1039.
 BENZIE, R. J., and COOKE, A. H., 1948 (in preparation).
 BLEANEY, B., and PENROSE, R. P., 1948a, *Proc. Phys. Soc.*, **60**, 83; 1948b, *Ibid.*, **60**, 395.
 BLOEMBERGEN, N., PURCELL, E. M., and POUND, R. V., 1948, *Phys. Rev.*, **73**, 679.
 BROER, L. J. F., 1943, *Physica*, **10**, 801.
 BROER, L. J. F., DIJKSTRA, L. J., and GORTER, C. J., 1943, *Physica*, **10**, 324.
 FRÖHLICH, H., 1946, *Nature, Lond.*, **157**, 478.
 GIAUQUE, W. F., and STOUT, J. W., 1941, *J. Amer. Chem. Soc.*, **63**, 714.
 GORTER, C. J., TEUNISSEN, P., and DIJKSTRA, L. J., 1938, *Physica*, **5**, 1013.
 DE HAAS, W. J., and DU PRÉ, F. K., 1938, *Physica*, **5**, 969.
 VAN DEN HANDEL, J., and SIEGERT, A., 1937, *Physica*, **4**, 871.
 HOWARD, J. B., 1935, *J. Chem. Phys.*, **3**, 813.
 KITTEL, C., and LUTTINGER, J. M., 1948, *Phys. Rev.*, **73**, 162.
 KRAMERS, H. A., 1930, *Proc. Amsterdam Acad. Sci.*, **33**, 959.
 KRISHNAN, K. S., and MOOKHERJI, A., 1938, *Phil. Trans. Roy. Soc. A*, **237**, 135.
 LAPORTE, O., 1928, *Z. Phys.*, **47**, 761.
 POLDER, D., 1942, *Physica*, **9**, 709.
 POWELL, H. M., and WELLS, A. F., 1935, *J. Chem. Soc.*, 359.
 PRYCE, M. H. L., 1948, *Nature, Lond.*, **162**, 539.
 SCHLAPP, R., and PENNEY, W. G., 1932, *Phys. Rev.*, **42**, 666.
 SIEGERT, A., 1936, *Physica*, **3**, 85; 1937, *Ibid.*, **4**, 138.
 VAN VLECK, J. H., 1939 a, *J. Chem. Phys.*, **7**, 61; 1939 b, *Le Magnétisme*, Vol. III (Strasbourg: Collection scientifique), p. 137; 1940, *Phys. Rev.*, **57**, 426.
 VAN VLECK, J. H., and WEISSKOPF, V. F., 1945, *Rev. Mod. Phys.*, **17**, 227.
 WEISS, P. R., WHITMER, C. A., TORREY, H. C., and JEN-SEN HSIANG, 1947, *Phys. Rev.*, **72**, 975.

Atmospheric Absorption of Millimetre Waves *

By H. R. L. LAMONT

Research Laboratories of The General Electric Company, Limited, Wembley

*MS. received 7 August 1948; read at Oxford Conference on Microwave Spectroscopy
23 July 1948*

ABSTRACT. The attenuation of electromagnetic waves in the region of 5 mm. wavelength due to resonance in the oxygen molecule was measured over atmospheric paths up to 2 km. long, using a fixed transmitter and a mobile superheterodyne receiver. For ten wavelengths between 6.34 and 4.48 mm. values were obtained which, after reduction to standard dry atmospheric conditions, range up to a maximum of 15.7 db/km. at the peak of the absorption curve. They agree substantially with the theory of Van Vleck, and give a value of line-breadth constant $\Delta\bar{\nu}$ of about 0.02 cm^{-1} , confirming estimates derived from other methods.

§ 1. INTRODUCTION

THE existence of a molecular resonance in oxygen giving rise to absorption of electromagnetic radiation in a wavelength region around 5 mm. has been predicted on theoretical grounds by Van Vleck (1947 a), who has discussed the theory at length. The absorption arises from the permanent magnetic moment associated with the oxygen molecule, transitions between closely-spaced components in the ground state giving resonances in the millimetre wave region. At atmospheric pressures the effect of collision broadening reduces individual resonances to a single broad band. An undetermined constant in the theory was the line-breadth constant $\Delta\bar{\nu}$, the wave-number half-width at half-intensity, which originally could only be estimated roughly.

Laboratory measurements confirming the effect have been made by Beringer (1946), who used a short length of waveguide filled with pure oxygen or a mixture of oxygen and nitrogen, and measured the attenuation constant at various wavelengths. His results gave a value for $\Delta\bar{\nu}$ between the limits 0.02 and 0.05 cm^{-1} .

The present paper describes experiments already briefly reported (Lamont 1948) carried out in the open air to measure the attenuation rate as it actually occurs in the atmosphere. The measurements were done on an R.A.F. aerodrome during the summer of 1947. Wavelengths in the region between 6.34 mm. and 4.48 mm. were used, with distances between transmitter and receiver up to 2.2 km. As will appear, the results provide satisfactory confirmation of the theory, and add weight to previous estimates of the line-breadth constant.

§ 2. APPARATUS

The equipment used was built for a general investigation of the propagation characteristics of millimetre waves, and some early work with it on propagation over sea has been reported (Lamont and Watson 1946). Briefly, the transmitter consists of a velocity-modulation oscillator feeding into a distorter crystal, in the form of a germanium crystal valve mounted in a cylindrical resonator, from

* Communication from the Staff of the Research Laboratories of The General Electric Company, Limited, Wembley, England.

which second harmonic power is extracted. This is mounted behind a 16-inch diameter aluminium paraboloidal reflector, which is fed from the front through a waveguide and small flare, the radiation being vertically polarized. Square-wave modulation at 1000 c/s. repetition rate is applied to the cathode screen.

The receiver is of superheterodyne type, with a germanium crystal mixer, and a velocity-modulation local oscillator operating at half the signal frequency. The aerial system is similar to that of the transmitter. A miniature I.F. amplifier of about 10 Mc/s. bandwidth is mounted directly below the mixer, and is connected by cable to an L.F. amplifier having a cathode-ray oscillograph and rectifier meter connected to its output. The receiver is remotely tuned by a Selsyn drive. The wavelength range was covered by three different valve types, two of which were made by E.M.I. Ltd. specially for this work.

Since the beam width of the aerial is small (about 1° between half-power points) each aerial carries a telescope aligned with the beam, and the aerials are then directed entirely by this optical means.

For these measurements the transmitter was on a fixed site, and the receiver was mobile, being mounted on a turntable on the roof of a small trailer, inside which was the measuring and calibrating equipment. Power was provided by a petrol-electric set in the towing vehicle, and the necessary voltage stabilization was incorporated in the power supply units.

§ 3. METHOD OF MEASUREMENT

If the transmitter and receiver are separated by a distance d , and the received signal is assumed to be free from ground reflections or other spurious echoes, then the received electric field E will be determined by the usual $1/d$ law with an additional attenuation α due to atmospheric absorption, so that $E = (E_0/d)e^{-\alpha d}$. From relative measurements of E made at a number of known distances a value of α can be determined.

For calibration of the receiver use is made of the well-established property that the I.F. output of a mixer crystal is proportional to the H.F. signal, so long as these are small compared with the amplitude of the local oscillator swing. A signal at 60 Mc/s., the mid-band frequency of the I.F. amplifier, is fed from a signal generator through resistance decoupling to the output side of the crystal mixer. This signal is modulated with a square wave of the same characteristics as the transmitter modulation, and is observed at the output of the receiver alternately with the signal received through the mixer. The change-over is done by means of two relays, one of which is a type with high-frequency contacts which form a switch in the coaxial cable from signal generator to mixer. The other relay is mounted at the side of the aerial mirror, and, through a push rod, operates two spring fingers which open or close over the end of the waveguide feed to the mirror. Thus operation of a foot switch effects instant change-over from received to calibrating signal, and the signal generator output can be rapidly adjusted to match the received signal in amplitude. Spot calibration of this type is necessary in order to make the measurements sufficiently independent of variations in the I.F. and L.F. gains over the comparatively long period required for a set of readings. The supply voltage and the output monitor reading of the signal generator were reset if necessary to their prescribed values during each reading.

Thus readings are dependent on the attenuator of the signal generator, and this has been calibrated by the National Physical Laboratory at the operating

frequency to an overall accuracy of ± 0.1 db. The accuracy of setting of the output monitor and the accuracy of matching of signals are both better than 0.1 db.

The transmitter is not monitored, other than by noting values of the rectified current in the distorter crystal. This is not in itself a good indication of power output, since the efficiency of frequency doubling is poor, and bears little relation to the driving power, but it does provide a check lest any variation in crystal characteristic should occur. The harmonic power output is of the order of $10 \mu\text{w.}$, for which level no satisfactory means of monitoring has yet been found. The frequency is determined by means of a cavity wavemeter loosely coupled to the waveguide connecting oscillator and distorter cavity, the accuracy being to within about 1 part in 1000.

§ 4. EXPERIMENTS

Measurements based on the relation quoted above assume free-space propagation conditions. An attempt was made first of all to find a site where such conditions would be evidently fulfilled, for example a deep valley with a road running along the top on one side, but nothing practicable was found.

The next best scheme was to use a flat piece of land with a tower on which the transmitter could be mounted. With this arrangement it will be seen from figure 1 that, although the receiver height may be small, the angle between direct and reflected rays at the receiver is dependent only on the transmitter height, and, therefore, with a sufficient transmitter height, a receiving aerial of narrow beam angle will discriminate between these rays.

A suitable site was found on an R.A.F. aerodrome, where permission was obtained to mount the transmitter on the flat roof of the Flying Control building and to use the perimeter tracks and runways for signal strength measurements. The transmitter height thus obtained was 46 feet.

After preliminary runs to try out the equipment, a number of changes were made, notably the addition of the rapid change-over from measured to calibrating signal described above, which improved both the speed and accuracy of the measurements. The tests included a check on the importance of differences between the modulation of transmitter and calibrating signal. It was found that the calibration was independent of wide changes in the repetition rate of one of the modulators, and was affected only in absolute level by changes in pulse width. Thus the two modulation waveforms must remain constant but need not be identical.

During the preliminary tests a number of sites were selected on the runways and perimeter tracks. They were all positions marked either by taxi lights or flare path lights, whose positions could be determined accurately from their representations on a map of the aerodrome having grid lines at 100-foot intervals. About 12 sites were used, varying in range from 0.121 km. to 2.193 km. The site nearest to the transmitter was the origin and terminal point of all runs, and the selection of sites for each run was made as a result of experience. It was usual to repeat all outward points on the inward run, but on occasions a circular tour was made.

If measured signals are to have their free-space value within ± 0.1 db., which was the accuracy aimed at, the amplitude of an interfering signal caused by ground reflection must have a value of less than 1.2% of the direct signal, that is, it must be at least 38 db. below the direct signal. It was found from the aerial polar

diagram that the gain is at least 20 db. below the maximum for all radiation received at angles greater than 1.4° from the axis. From the formula of figure 1, $\theta = 2h_1/d$ (h_1 being the transmitter height), θ will be greater than 1.4° for values of d less than 1.1 km. Thus ground reflections should give no trouble within this range if the remaining 18 db. at least is lost at the point of reflection, which corresponds to a reflection coefficient of about 0.1. No direct measurement of this reflection coefficient was attempted, but an estimate of its value can be obtained from the work of Ford and Oliver (1946) at a wavelength of 9 cm.

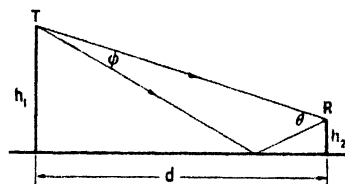


Figure 1. Angular relations between direct and reflected waves.

If $d \gg h_1$, then $\theta = 2h_1/d$, $\phi = 2h_2/d$.

Their measurements show that a covering of vegetation on level ground entirely eliminates specular reflection, and with vegetation about 40 cm. high a reflection coefficient of 0.04 was obtained with vertical polarization and an angle of incidence of 68° . Rough ground covered with long thick grass gave a value of 0.08 for the same conditions. Scaled to 5 mm. wavelength, these surfaces are probably rather smoother than the moderately short grass of the aerodrome, and, although the reflection coefficient will increase as grazing incidence is approached, it seems that a maximum value of 0.1 under the conditions of the present experiments is not unreasonable.

The site at a range of 1.047 km. was the most distant site at which signals could be measured accurately for the wavelengths where the attenuation was high. For the lower attenuations the range was extended to a maximum of 2.193 km. These latter are beyond the range established above for which interference from ground reflections should be negligible, and in fact reflections were found to exist at the more distant sites. The interference was not regular, and, presumably on account of a complex interference field from a reflecting surface which in places was far from plane, sites differing in position by only a few yards sometimes gave quite markedly different signal levels. Most of these points have been included in the calculations since it is considered that the net effect of in-phase and out-of-phase reflections averaged over several sites will be small.

Ten different wavelengths were used, varying between 6.34 and 4.48 mm. The shortest wavelength was obtained, for the transmitter, as the third harmonic of a 13.44 mm. oscillation. In this case the beating oscillator of the receiver (kindly lent by the Clarendon Laboratory, Oxford) operated at 8.96 mm. wavelength. With the exception of the shortest wavelength, the experiments were carried out between 1st August and 5th September 1947, a period of fine warm weather. The experiments on the final wavelength were done during a cold spell at the end of November. The results are calculated from a total of thirty-five runs, each comprising, on the average, seven readings.

Values of air temperature, pressure and humidity were obtained from readings taken hourly by the meteorological staff of the aerodrome in the vicinity of the transmitter.

§ 5. RESULTS

The field strength at a distance d is given, as already stated, by

$$E = (E_0/d)e^{-\alpha d} \dots\dots (1)$$

Measurements are obtained in the form of values of E expressed in decibels relative to an arbitrary level for known values of d . Putting the formula, therefore, in logarithmic form, we get

$$20 \log E + 20 \log d = 20 \log E_0 - \alpha d, \quad \dots\dots(2)$$

where αd is now expressed in decibels instead of nepers as in (1). On replacing E by its decibel level N this becomes

$$N + 20 \log d = -\alpha d + N_0. \quad \dots\dots(3)$$

Thus if values of $N + 20 \log d$ are plotted as ordinates y against d as abscissa the points should lie on a straight line whose slope is $-\alpha$. Figure 2 is a typical set of readings for a single run, showing their deviations from this linear relationship.

The best straight line through actual points subject to small errors is most accurately determined by the Method of Least Squares. With the values of d taken as exact, this gives for α , the most probable value of α (Birge 1932),

$$\bar{\alpha} = \frac{\Sigma y \Sigma d - s \Sigma y d}{s \Sigma d^2 - (\Sigma d)^2}, \quad \dots\dots(4)$$

where s is the number of points, and summations extend over these points. To calculate the probable error of $\bar{\alpha}$ we must also find the mean value $\bar{N}_0$ of the constant N_0 , which is given by

$$\bar{N}_0 = \frac{\Sigma d^2 \Sigma y - \Sigma d \Sigma y d}{s \Sigma d^2 - (\Sigma d)^2}. \quad \dots\dots(5)$$

When values given by (4) and (5) are substituted in the s equations (3), a set of residuals $v_1, v_2, \dots, v_s$ is obtained, from which the probable error in $\bar{\alpha}$ is derived as

$$\epsilon = 0.67 \left[\frac{\Sigma v^2}{s-2} \frac{s}{s \Sigma d^2 - (\Sigma d)^2} \right]^{\frac{1}{2}}. \quad \dots\dots(6)$$

The values of $\bar{\alpha} \pm \epsilon$ thus obtained are shown in the third column of the table.

Run No.	λ	$\bar{\alpha} \pm \epsilon$	$\bar{\alpha}_c$	$\bar{\alpha}_m$	P.E.	Run No.	λ	$\bar{\alpha} \pm \epsilon$	$\bar{\alpha}_c$	$\bar{\alpha}_m$	P.E.
23	6.34	0.2 $\pm$ 0.15	0.05	0.05	0.15	44	5.13	15.49 $\pm$ 0.4	15.7	15.7	0.2
25	5.76	1.24 $\pm$ 0.35	1.08	1.0	0.1	45	5.13	15.71 $\pm$ 0.3	15.9		
26	5.76	0.98 $\pm$ 0.1	0.83			46	5.13	15.26 $\pm$ 0.2	15.5		
27	5.76	1.37 $\pm$ 0.2	1.24			47	5.13	15.44 $\pm$ 0.5	15.9		
28	5.76	1.22 $\pm$ 0.1	1.08			48	5.04	14.28 $\pm$ 0.2	14.6	14.5	0.15
29	5.76	1.48 $\pm$ 0.2	1.33			49	5.04	13.96 $\pm$ 0.3	14.2		
31	5.60	2.19 $\pm$ 0.4	2.07	1.8	0.2	50	5.04	14.50 $\pm$ 0.3	14.7		
34	5.60	1.80 $\pm$ 0.2	1.71			51	5.10	14.87 $\pm$ 0.3	15.15	13.9	0.1
35	5.28	9.94 $\pm$ 0.3	9.9	10.2	0.1	52	5.10	13.99 $\pm$ 0.2	14.35		
36	5.28	9.90 $\pm$ 0.1	10.3			53	5.10	14.26 $\pm$ 0.4	14.42		
37	5.28	9.75 $\pm$ 0.35	10.1			54	5.10	13.35 $\pm$ 0.15	13.6		
39	5.28	9.42 $\pm$ 0.35	9.7			55	5.10	13.41 $\pm$ 0.2	13.6		
40	4.96	14.60 $\pm$ 0.4	15.2	14.7	0.2	56	5.10	13.96 $\pm$ 0.25	14.1		
41	4.96	14.30 $\pm$ 0.3	14.8			57	5.19	12.97 $\pm$ 0.1	12.8	12.7	0.05
42	4.96	14.17 $\pm$ 0.1	14.6			58	5.19	12.75 $\pm$ 0.1	12.7		
43	4.96	14.22 $\pm$ 0.25	14.7			59	5.19	12.92 $\pm$ 0.1	12.8		
						60	4.48	0.75 $\pm$ 0.3	0.57	0.4	0.1
						61	4.48	0.47 $\pm$ 0.1	0.31		
						62	4.48	1.04 $\pm$ 0.5	0.84		

λ =wavelength (mm.); $\bar{\alpha} \pm \epsilon$ =measured attenuation (db/km.); $\bar{\alpha}_c$ =corrected attenuation (db/km.); $\bar{\alpha}_m$ =mean $\bar{\alpha}_c$ for each wavelength; P.E.=probable error ($\pm$ db/km.).

These are values measured under differing atmospheric conditions and have now to be reduced to correspond to a standard atmosphere. The first correction is for water vapour. In his analysis of absorption by water vapour Van Vleck (1947 b) treats the attenuation in two parts: (a) that due to the absorption line at 1.35 cm. wavelength, and (b) the combined residual effect of other lines of much shorter wavelength. The theoretical values of these for the range of interest are shown as curves (a) and (b) of figure 3 for air containing 1 gm. of water per

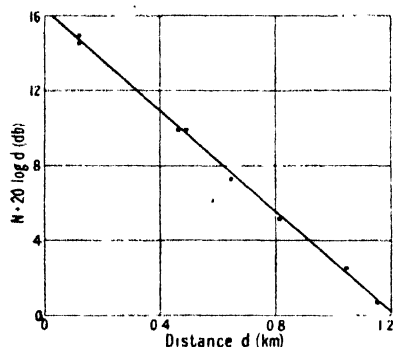


Figure 2. Typical set of observations plotted to give the attenuation constant as the gradient of a straight line.

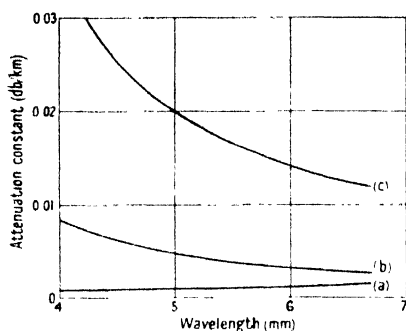


Figure 3. Attenuation due to water vapour, per gramme H_2O per m^3 , at $293^\circ K.$, 760 mm. Hg.

cubic metre. From a discussion of the experimental results of several workers at wavelengths around 1.3 cm. Van Vleck concludes that the theoretical values for (a) are correct, but for (b) should be increased by a factor of about 4. The sum of the two parts, with this modification, is given in figure 3, curve (c), and these values are used for the reduction of the present data. An attenuation appropriate to the actual concentration $G \text{ gm}/m^3$ of water vapour in the air, determined from the humidity, is subtracted from each measured attenuation rate. The values of G vary between 5.1 and 16 gm/m^3 , corresponding to attenuation rates between 0.13 and 0.31 db/km .

The remainder must now be corrected for temperature and pressure, the standards used being $293^\circ K.$ and 760 mm. Hg. The theoretical formula given by Van Vleck (1947 a) for the attenuation due to oxygen is of the form $\alpha = FN/T$, where N is the number of molecules of oxygen per unit volume, T the absolute

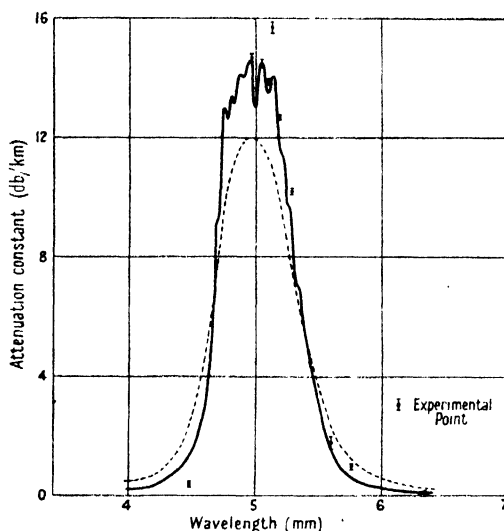


Figure 4. Measured values of attenuation due to oxygen, at $293^\circ K.$, 760 mm. Hg.

Theoretical curves $\left\{ \begin{array}{l} \text{— } \Delta\bar{\nu} = 0.02 \text{ cm}^{-1}. \\ \text{--- } \Delta\bar{\nu} = 0.05 \text{ cm}^{-1}. \end{array} \right.$

temperature, and F a complex function of individual resonance frequencies. It is known (Beringer 1946) that the collision cross-section is substantially the same for O_2-O_2 and O_2-N_2 impacts. If we assume in addition that O_2-H_2O impacts are equally effective, then the attenuation will be proportional to the total number of molecules per unit volume and hence to the total pressure p . This is a reasonable assumption since the kinetic theory collision diameters for O_2-H_2O and O_2-N_2 impacts are respectively 4.1 Å. and 3.7 Å.; other work in this field has shown that collision diameters are little different from those calculated by kinetic theory. As N varies inversely with T , the dependence on temperature is approximately as T^{-2} , F being relatively independent of temperature. In these measurements the temperature and pressure ranges were 290–301° K. and 755–763 mm. Hg, with the exception of the figures for the shortest wavelength, which were 278° K. and 745 mm. Hg. The maximum combined correction for temperature and pressure was 8%.

Finally the values are corrected to correspond to a dry atmosphere containing 21% oxygen molecules as used by Van Vleck. This requires multiplication by a factor $p/(p-p')$ where p' is the partial pressure of the water vapour, and is related to the water vapour concentration G by the formula $p' \text{ (mm.)} = 0.0035 GT$. The amount of this correction varies between 1% and 2%.

The finally corrected values of $\bar{\alpha}$ for each run are given in the fourth column of the table. The average for each wavelength group is then taken, each reading being weighted according to its probable error. The final results with their probable errors, based on internal consistency, are shown in the last two columns, and in figure 4 these values are plotted against wavelength.

A further uncertainty of ± 0.1 db/km. should be added to each result to allow for signal generator calibration uncertainty.

§ 6. DISCUSSION OF RESULTS

Figure 2 shows points diverging by amounts up to about 0.3 db. from the linear relationship, and this is typical of the results obtained. It is considered that part of this variability is due to interference from signals arriving by other than the direct path, even on the shorter ranges. This could be improved by re-design of the receiving aerial system to sharpen the lower half of the polar diagram, or preferably by using a higher transmitting site. Some of the variation must also be attributed to slow changes in the transmitter output and to a certain amount of instability in the distorter and mixer crystals, which, requiring extremely light contact and having little whisker length to allow flexibility, cannot be so robust as the capsule crystals familiar at longer wavelengths.

As regards dependence on external conditions, temperature is the most important, and a change of 1°C., for example, would produce a change of about 0.1 db/km. for the larger values of attenuation. There will be variations in temperature over the path, characterized by horizontal and vertical gradients and irregular short-period variations. These have been discussed by Sheppard (1946), and from his paper it is clear that the difference between the temperature at the transmitter and receiver heights may easily vary by 1°C. during the day, and short-period variations at a fixed point up to $\pm 1^\circ\text{C.}$ are usual. These uncontrollable variations are probably the limiting factor in measurements of this type.

In figure 4 the solid curve is that calculated by Van Vleck for a line-breadth constant $\Delta\bar{\nu}$ of 0.02 cm^{-1} , and the values are obviously in good agreement with

the curve. The spectroscopic determinations of the energy intervals from which the curve is calculated have not sufficient exactitude to allow details of its incipient fine structure to be plotted accurately, but the curve is a good approximation. That given here differs very slightly from Van Vleck's curve, in that spectroscopic values of higher accuracy by Babcock, quoted but not used by Van Vleck, have been used. The dotted curve in figure 4 is for $\Delta\bar{\nu}=0.05\text{ cm}^{-1}$, a width too great to show any individual lines, and it is evident that the measurements do not support this value. The value of 15.7 obtained for α at 5.13 mm. wavelength, which is well above the curve, may indicate a line-breadth constant rather less than 0.02 cm^{-1} , with a consequent increase in the amount of structure appearing.

Beringer's measurements, made in the laboratory over a similar range of wavelengths, are in general agreement with the values given here, but show a greater spread. Considering Beringer's work and some indirect evidence, Van Vleck deduced that about 0.02 cm^{-1} was the best estimate of $\Delta\bar{\nu}$. The present work substantiates this. Here, as also in Beringer's measurements, it would appear that the curve should be rather wider on its long-wavelength side than the existing theory would indicate. On the other hand the single point on the skirt at the short-wavelength end indicates the opposite effect. In his paper Van Vleck remarks that "it should not be regarded as highly surprising if the detailed shape of the absorption curve is somewhat different from that calculated by means of our formulas". On the whole the points in figure 4 suggest that the curve should be moved slightly to the right along the wavelength axis, but there would appear to be no warrant for this on a theoretical basis. For the moment it can be said that the present results confirm the theory by a very direct method and add weight to the previous estimate of the line-breadth constant. The value of 0.02 cm^{-1} for $\Delta\bar{\nu}$ gives an effective collision diameter of 2.8 Å., which is not far from the collision diameter value of 3.6 Å. given for oxygen by kinetic theory.

REFERENCES

- BERINGER, R., 1946, *Phys. Rev.*, **70**, 53.
 BIRGE, R. T., 1932, *Phys. Rev.*, **40**, 207.
 FORD, L. H., and OLIVER, R., 1946, *Proc. Phys. Soc.*, **58**, 265.
 LAMONT, H. R. L., 1948, *Phys. Rev.*, **74**, 353.
 LAMONT, H. R. L., and WATSON, A. G. D., 1946, *Nature, Lond.*, **158**, 943.
 SHEPPARD, P. A., 1946, *Meteorological Factors in Radio-Wave Propagation* (London: The Physical Society), p. 37.
 VAN VLECK, J. H., 1947 a, *Phys. Rev.*, **71**, 413; 1947 b, *Ibid.*, **71**, 425.

Note on Nuclear Magnetic Resonance

By E. E. SCHNEIDER

King's College, Newcastle-on-Tyne (University of Durham)

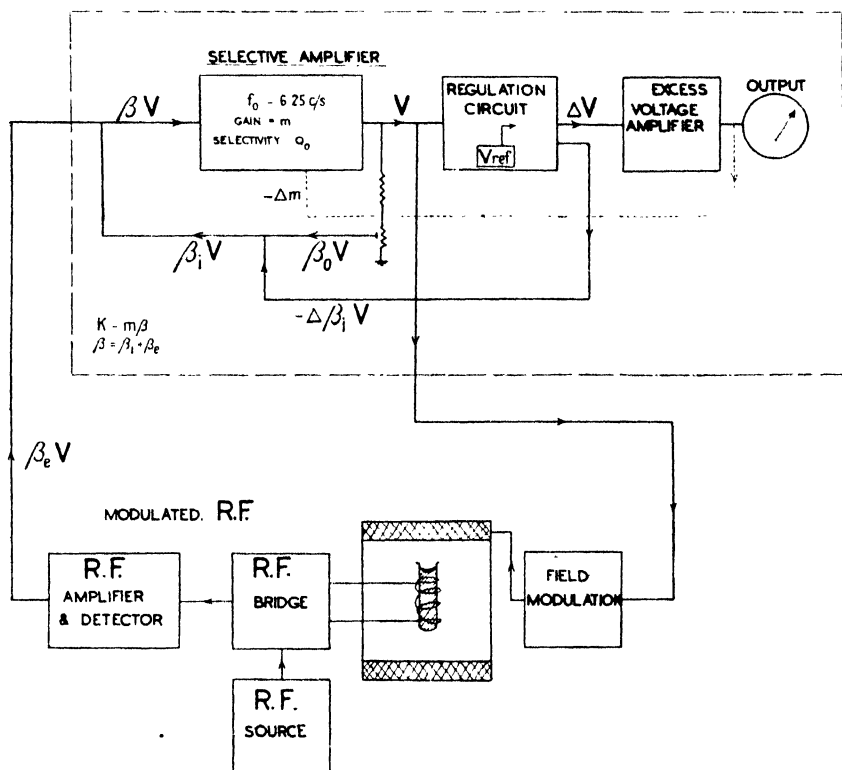
*MS. received 6 August 1948; read at Oxford Conference
on Microwave Spectroscopy 23 July 1948*

I SHOULD like to report briefly on a general method of physical measurement which seems ideally suited for the detection of very weak nuclear magnetic resonances, and which promises to be very useful for the absolute measurement of relaxation times.

The main idea of the method is to increase the discrimination against noise by utilizing the extreme frequency selectivity of a continuously oscillating system maintained by non-linear regeneration. According to the classical work of Appleton and van der Pol the inherent damping of the original resonator must be overcompensated by the regenerative action so that the effective damping λ in the maintained state becomes negative and the reproduction factor k , giving the ratio of the regenerated to the original oscillating quantity, becomes greater than unity. For given amplitude the values of $k - 1$ and λ decrease with decreasing non-linearity. The selectivity of the system with respect to driving forces of different frequencies around the oscillation frequency is essentially given by $|\lambda|$ and can, therefore, be very large if the non-linearity is small.

Such a non-linear system can serve for the measurement of a physical effect if the effect itself can be made to provide an additional regenerative action. This external regeneration will add up to the internal regeneration already present and so change the oscillation amplitude. More convenient is the use of a regulation method to measure the external effect by the compensating change of the internal regeneration required to keep the amplitude constant.

The diagram shows the application of this method to the study of nuclear



magnetic resonance. The oscillating system consists of a selective very low frequency amplifier with positive feedback providing the main internal regeneration. The oscillation amplitude is regulated by a negative feedback contribution resulting from a deviation of the output from a reference value. This deviation or excess voltage indicates the magnitude of the effect under observation since it is a measure of the change in β_1 compensating the external feedback. The

oscillator output is used to modulate the magnetic field of the otherwise conventional nuclear resonance set-up. With the mean magnetic field adjusted near resonance the radio frequency is, therefore, modulated at the low oscillator frequency and the detected output from the R.F. receiver forms a feedback voltage which combines vectorially with the internal feedback voltage. Its phase depends on the position on the magnetic resonance curve of the mean magnetic field, on the modulation index, and on the relaxation time and the R.F. intensity.

With a system of this type values of $k - 1$ of the order of 10^{-3} have been obtained and changes of k of less than 10^{-3} have been detected. Since $\Delta k = \Delta \beta m$, $V_1 = \Delta \beta V = \Delta k V / m$, and the gain $m = 5 \times 10^4$ and the output voltage $V = 10$ v., this means that voltage changes at the low-frequency oscillator input of smaller than 2×10^{-7} v. are observable across an input resistance greater than 10^8 ohm. When using the R.F. arrangement, providing an additional gain of 10^3 , observable deflections of the final output meter have been obtained by a modulation component of less than 10^{-9} v. across the R.F. input resistance of 3×10^4 ohm, i.e. by an available input power of less than 10^{-23} watt. This high sensitivity in the presence of thermal noise could be interpreted in terms of a bandwidth of less than $1/400$ c/s. It is very doubtful, however, whether the concept of a bandwidth for noise has any sense in a non-linear system. It is certainly not correct to consider the reciprocal of the bandwidth as the response time which can here not be defined as the time constant of a simple exponential decay or build up of the response. The time actually taken by the system to respond to the small input voltage changes just quoted is of the order of several seconds and not minutes as would follow from the bandwidth.

The possibility of obtaining very high sensitivities, very high "signal to noise ratios", without a sacrifice in the speed of observation is the most important achievement of the new method and a direct consequence of its essential non-linearity.

CORRIGENDA

"A Radio-sonde Method for Atmospheric Potential Gradient Measurements", by R. E. BELIN (*Proc. Phys. Soc.*, 1948, **60**, 381).

I wish to point out that the work as described under the above title was carried out at the Auckland University College, New Zealand (and not at the address stated in the paper) under the supervision of Professor P. W. Burbidge and Dr. K. Kreielsheimer.

The latter has kindly pointed out an error in equation (3), p. 385, which should read $i = i_0(P_0/P)^{1/6}$, and a better explanation of figure 6, viz., the figure suggests a negative distribution of charge residing both at the top and the bottom, with a positive distribution near the centre of the cumulo-nimbus cloud, while the fracto-stratus formation may have a positive charge at the bottom and negative at the top.

It should also have been emphasized that the formulae used in the paper are only approximate.

"Travelling Wave Linear Accelerators", by R. B. R-SHERSBY-HARVIE (*Proc. Phys. Soc.* 1948, **61**, 255).

In figure 9 the ordinate scale is too great by a factor of $\sqrt{3}$.

ABSTRACTS OF PAPERS READ AT OXFORD CONFERENCE

23rd and 24th July 1948

Spectroscopy at Radio and Radar Frequencies

By C. J. GORTER

Leiden

Brief opening remarks (*see this issue*, p. 541),

The Theory of Line Breadths in Microwave and Radio Frequency Spectra

By J. H. VAN VLECK

Harvard

Three mechanisms for line broadening may be distinguished : (a) collision broadening of the Lorentz type, applicable to the microwave spectra of gases, (b) collision modulation of intramolecular nuclear spin-spin coupling, whose theory has been developed by Bloembergen, Pound and Purcell, and (c) the adiabatic broadening by dipolar interaction of the magnetic resonance of regular crystals, devoid of molecular rotation. The present paper is concerned primarily with (c), the simplest process of the three. The eigenvalue problem is still too complicated to permit calculation of the detailed line profile, but the computed mean square and mean fourth power frequency deviations agree excellently with Pake's measurements on the magnetic resonance of the F nucleus in CaF_2 , both as regards absolute magnitude and directional dependence. The line contour in CaF_2 is almost gaussian. On the other hand, the observed half-widths of microwave absorption lines arising from electronic spin in paramagnetic salts are considerably smaller than calculated from the theoretical mean square if a Gauss distribution is assumed. The discrepancy is shown to be caused by the phenomenon of "exchange narrowing", wherein the fluctuations produced by exchange forces spoil the coherence and efficiency of the dipolar coupling. (*To be published in Physical Review*, 1948.)

Experimental Investigations of Paramagnetic Resonance Lines

By J. H. E. GRIFFITHS

Oxford

Paramagnetic resonance in $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$ has been investigated at wavelengths of 3.04 cm., 1.23 cm. and 0.85 cm. Two features of considerable interest are observed, first, that under certain conditions, the absorption line is considerably narrower than would be expected from the normal magnetic interactions of the ions, and second, that at the longer wavelengths only one line is observed whereas, at certain orientations of the crystal in the external magnetic field, two lines should be observed. The two lines are clearly resolved at 0.85 cm. wavelength. A qualitative explanation of these results is given in terms of exchange interaction. These results are also compared with those for $\text{CuSO}_4 \cdot \text{K}_2\text{SO}_4 \cdot 6\text{H}_2\text{O}$ in which the exchange interaction appears to be considerably weaker. (*See D. M. S. Bagguley and J. H. E. Griffiths, Nature, Lond.*, 1948, **162**, 538.)

Theory of the Line Breadth in Copper Sulphate

By M. H. L. PRYCE

Oxford

The observations on copper sulphate ($\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$) suggest that the rapid exchange between neighbouring ions so reduces the effect of the internal magnetic field that the breadth due to magnetic interaction of the ions is only a minor part of the observed breadth. The main breadth can be explained as arising from exchange between crystallographically dissimilar ions.

In a crystal in which all the paramagnetic ions are similar, exchange cannot, in the absence of magnetic interaction, have any broadening effect on the resonance. In copper sulphate, however, there are two kinds of ions, differing in the orientation of the atoms in their neighbourhood. Exchange between the two leads to a broadening. A quantitative calculation of the effect is possible for low frequencies (up to about 10 000 Mc/s.). It shows that for certain orientations of the magnetic field the breadth increases as the square of the frequency. The dependence on orientation of the field is given by a formula containing three undetermined parameters. (*See Nature, Lond.*, 1948, **162**, 539.)

Paramagnetic Resonance at Low Temperatures

By B. BLEANEY

Oxford

A general survey has been made of salts of the iron group by the method of paramagnetic resonance to establish the conditions under which absorption lines are detectable. At room temperature and 3 cm. wavelength lines are observed in salts of $3d^3 \text{Cr}^{+++}$, $3d^4 \text{Cr}^{++}$, $3d^5 \text{Fe}^{+++}$ and Mn^{++} , and $3d^9 \text{Cu}^{++}$. In $3d^1 \text{Ti}^{+++}$, $3d^6 \text{Fe}^{++}$, $3d^7 \text{Co}^{++}$ the spin-lattice relaxation time is too short at room temperature but resonance can be observed at sufficiently low temperatures. Short wavelengths are required for $3d^8 \text{Ni}^{++}$ (and also Fe^{++}) because of the large separation of the spin levels in zero magnetic field. These effects are associated with the splitting of the orbital levels caused by the electric field of the crystalline lattice.

Single crystals of four chrome alums have been studied. At room temperature the spectra can be interpreted assuming the lattice produces a small trigonal electric field superimposed on a cubic field. The spin levels consist then of two doublets with a separation varying from 0.12 cm^{-1} for the potassium alum to 0.16_3 cm^{-1} for the rubidium alum. As the temperature is reduced the splittings decrease slightly in rubidium and caesium alums while in potassium and ammonium alums a considerable initial decrease is followed by the appearance of different types of spectra at liquid oxygen temperatures. Each of these two alums then exhibits two splittings, presumably associated with different ions in unit cell. (*See this issue*, pp. 542, 551.)

Pressure Broadening of the Ammonia Spectrum

By B. BLEANEY

Oxford

At centimetre wavelengths the dominant contribution to the width of gaseous spectral lines is due to collision broadening. It is thus possible to study collision broadening over a wide range of pressures, and the absorption due to the strong inversion spectrum of ammonia near 1 cm. wavelength has been measured over the range 6 atmospheres down to a few hundredths of a millimetre. The main features are :

(a) From measurements of the line widths at 0.5 mm. Hg, the collision diameter is found to vary from 9 Å. to 14 Å., which is considerably greater than the kinetic theory diameter, 4 Å. This, and the regular change in width with the rotational state of the molecule are explained in terms of dipole-dipole interaction. The cross-sections for collisions with non-polar gases are approximately equal to the kinetic theory values.

(b) At pressures up to 10 cm. Hg, the absorption curve agrees closely with that calculated on the assumption that the line width varies linearly with the pressure, but at higher pressures the centre of the absorption shifts to lower frequencies and the line width increases less rapidly than the pressure. Above two atmospheres the inversion frequency becomes zero and the collision diameter tends to a new constant value of 8 Å.

(c) At pressures below a few tenths of a millimetre a "saturation" effect is observed, since the collision frequency becomes too low to counteract the disturbance of thermal equilibrium produced by the radiation. Comparison with theory shows that the thermal relaxation time is equal to the mean time between collisions. (See Bleaney, B., and Penrose, R. P., *Proc. Phys. Soc.*, 1948, **60**, 83.)

Microwave Absorption in Gases

By M. H. L. PRYCE

Oxford

In microwave spectroscopy one can use coherent incident radiation which is monochromatic to a high degree, and the breadth of absorption lines in gases is due either to molecular collisions or to the intensity of the incident radiation, Doppler effect and natural line breadth being negligible. Lorentz's classical theory of pressure broadening needs refining in two interesting respects. For very broad lines Lorentz's hypothesis that collisions leave the molecular oscillators with random phase needs to be modified. This has been done by Van Vleck and Weisskopf, and by Fröhlich, but the experimental material suggests that a further refinement is required, in which the collision process is considered in greater detail.

For high intensities of the incident radiation the Boltzmann distribution of the energy levels between which the microwave transitions are taking place is disturbed, giving rise to a broadening. This is a phenomenon not found in optical spectra. It is essentially a quantum phenomenon, and can be calculated by a simple extension of the usual absorption calculations. This theory will be discussed. (*To be published in Proc. Phys. Soc.*)

Atmospheric Absorption of Millimetre Waves

By H. R. L. LAMONT

General Electric Co.

The paper describes measurements of the attenuation, due to molecular absorption in oxygen, of electromagnetic radiation in the region of 5 mm. wavelength. The attenuation was measured as it occurs in the atmosphere, using paths up to 2 km. long over flat ground, with a fixed elevated transmitter and a mobile superheterodyne receiver. The relative signal strength at any point was determined by direct comparison with a 60 Mc/sec. signal generator having an accurately calibrated attenuator. Ten wavelengths between 6.34 and 4.48 mm. were used, making a total of thirty-five runs, each comprising about seven readings.

The mean attenuation derived from each set of readings was corrected for water vapour attenuation, and then reduced to a value appropriate to an atmosphere at 20° C. and 760 mm. Hg. The final attenuation values range up to a maximum of 15.7 db/km., with probable errors between 0.1 and 0.2 db/km. They are in good agreement with Van Vleck's theoretical curve for a line-breadth constant $\Delta\nu/c$ of 0.02 cm⁻¹, in confirmation of estimates of the line-breadth derived from other methods. (*See this issue*, p. 562.)

Nuclear Effects in Microwave Absorption in Molecules

By A. G. HILL

Massachusetts Institute of Technology

Some twenty-five lines of the microwave spectrum of the linear molecule OCS_e have been observed by Strandberg and others at M.I.T. Lines associated with the six stable isotopes of selenium and with C¹², C¹³ and O¹⁶ have been measured in the ground vibrational state. In addition some of the stronger lines have been observed for the symmetric vibrational and bending states. Most of the observations have been made for the J=2→3 transition, lying in the region near 25 000 Mc/s., but in addition a few lines have been observed for the J=5→6 and J=6→7 transitions. There is no evidence of quadrupole coupling for the Se⁷⁷ isotope greater than 0.5 Mc/s., and from a standpoint of reasonableness one concludes that the spin of Se⁷⁷ is $\frac{1}{2}$. Mass differences for the Se isotopes have been obtained assuming the values of Se⁷⁴ and Se⁸². Stark effect measurements enable one to calculate the dipole moment. Measurements of relative intensities for the ground state and for the bending state enable one to predict a vibrational line at 474 cm⁻¹. Lord and McDonald have subsequently detected a transition at 464 cm⁻¹. Relative intensity measurements are in very good agreement with known data on relative isotopic abundance. (*To be published in Physical Review or Proc. Phys. Soc.*)

Magnetic Relaxation of Nuclei

By N. BLOEMBERGEN

Leiden

Experiments on nuclear magnetic resonance in crystals of CaF₂ and KAl(SO₄)₂·12H₂O show that relaxation times are at least 10⁵ times shorter than predicted by Waller's theory, and are reduced strongly by small amounts of paramagnetic impurities. The dependence of the relaxation time on temperature has been investigated between 300° K. and 1° K. At the lowest temperature relaxation times as long as half an hour have been observed.

A new theory of the heat contact between nuclear spins and lattice vibrations has been developed. The energy absorbed by the spins is transported toward the impurities by a diffusion process. The impurities act as sinks, through which the energy is transferred to the lattice. If the diffusion is the limiting factor in the energy transfer, the relaxation time is independent of temperature, and it is a decreasing function, if the impurities are the bottleneck. The detailed behaviour will be very sensitive to the nature and amount of the paramagnetic impurity. The theoretical results agree, at least in order of magnitude, with the experimental data.

Experiments on Nuclear Resonance at Low Temperature

By B. V. ROLLIN

Oxford

Further measurements on nuclear paramagnetism at low temperatures have been made using the magnetic resonance absorption method. Experiments have been made on single crystals of calcium fluoride at temperatures between 20° K. and 0.2° K. in order to determine the effect of temperature, magnetic field strength and paramagnetic impurities on the spin-lattice equilibrium time of the fluorine nuclei.

Nuclear resonance has been observed in copper and aluminium at 1.1° K.

Measurements have been made on solid hydrogen at temperatures down to 1.1° K. A splitting of the resonance is observed at the lowest temperatures. (*To be published in Rep. Prog. Phys.*, Vol. XII, 1949.)

Effects of Nuclear Electric Quadrupole Moments in the Radio-frequency Spectra of Solids and Liquids

By R. V. POUND

Harvard

Several cases have been reported of nuclear magnetic resonance absorption in liquids, in which a fluctuating interaction between the nuclear electric quadrupole moment and the gradient of the electric field at the nuclei shortens the relaxation time and broadens the resonance line. Such an effect is very pronounced in the resonances of bromine, iodine and gallium in solutions. In crystalline solids of lower than cubic symmetry the absorption line should have a structure with a multiplicity depending on the spin of the nucleus. Evidence of such an effect was found in the Li^7 resonance in powdered Li_2CO_3 and $\text{Li}_2\text{SO}_4 \cdot \text{H}_2\text{O}$, and a definite triplet line was observed in a properly oriented single crystal of $\text{Li}_2\text{SO}_4 \cdot \text{H}_2\text{O}$. In a corundum crystal, the line from Al^{27} was found to have the five components characteristic of a spin of $5/2$ and the splitting was not very small compared to the centre frequency used. Accordingly, the first order terms in the perturbation energy were not sufficient fully to describe the interaction. Excellent agreement was found between calculations of the interaction carried to the third order and the experimental observations of the dependence of the line frequencies on the orientation of the crystal in the magnetic field. (*To be published in Proc. Phys. Soc.*)

Measurements of Relaxation Times

By L. E. DRAIN

Oxford

This direct method of determining nuclear spin-lattice relaxation times which are comparable with the period of the modulating field is based on Bloch's nuclear induction experiment. As the magnetic field is varied through resonance, the sizes of the signals on the forward and backward sweeps, depend on how long is spent above and below resonance and upon the relative magnitudes of the relaxation time and the period of sweep.

Under certain conditions, a simple relation holds between the size of the signal, its position on the trace and the relaxation time. A quick and convenient procedure is to observe the point at which one signal disappears. This gives the relaxation time directly. (*To be published in Proc. Phys. Soc.*)

Investigations of Ferromagnetism in the Microwave Region†

By J. H. E. GRIFFITHS

Oxford

Measurements of ferromagnetic resonance at 1.23 cm. wavelength in thin films of nickel are described. The films were obtained by evaporating nickel on to mica. The results may be expressed in terms of the equation $g\beta(BH)^{1/2} = h\nu$ obtained by Kittel, where β is the Bohr magneton, B and H the magnetic induction and field strength, ν the frequency and $g=2$ if the resonance is due to electron spin only. Previously reported measurements on bulk material have given values for g of about 2.15. Thin nickel films ($< 0.5\mu$ thick) give $g = 2.00 \pm 0.02$ while films greater than 0.8μ thick give $g = 2.14 \pm 0.02$ in agreement with the values for the bulk material. Intermediate values are obtained in the transition region. (*To be published in Nature, Lond.*)

Microwave Resonance Absorption in Ferromagnetic Materials

By C. KITTEL

Bell Telephone Laboratories

The ferromagnetic resonance absorption effect was discovered by Griffiths, who observed that the energy loss in a microwave cavity containing ferromagnetic material shows a resonance type behaviour as a function of the static magnetic field intensity applied to the cavity. The effect basically is the result of electron spin resonance absorption, and it has been shown by the present author that the resonance relation between frequency and applied field is $\omega = g(e/mc)(BH)^{\frac{1}{2}}$ for the particular case of a plane specimen with the applied field parallel to the plane. The relation may be derived classically, or else, on quantum mechanics using the Hamiltonian $H = -\vec{M} \cdot \vec{H} + 2\pi M^2\gamma$. The resonance relation depends, according to theory, on the specimen shape, the field orientation, and, in the case of a single crystal, on the orientation of the crystal axes.

These predictions of the theory are confirmed in several cases by the measurements of Yager, Kip and Arnold, and Hewitt. In several materials it appears that the g -value is in the neighbourhood of 2.2: this conclusion is particularly well confirmed in the case of Superalloy. There is a special interest in Heusler alloy, but preliminary measurements by Yager do not give a conclusive g -value because of the width of the resonance. (See *Physical Review*, 1948, 73, 155.)

Microwave Absorption in Metals at Low Temperatures

By J. C. SLATER

Massachusetts Institute of Technology

Measurements of the surface resistance and reactance of lead, tin and copper, at a wavelength of 1.25 cm., and temperatures down to about 2° K., have been carried out by a group at M.I.T. using a resonant cavity method, and the programme contemplates measurements on a number of other metals. The results are qualitatively similar to Pippard's measurements at longer wavelength. The decrease in bulk resistivity in going to very low temperatures is not accompanied by a corresponding decrease in surface resistance, as would be calculated by the classical theory of skin effect. Furthermore, below the superconducting transition, the surface resistance does not go to zero, but decreases at first rapidly, then slowly, remaining finite at 2° K. These results are compared with the theories of London, Pippard, and Sondheimer and Reuter, based on the assumption that the electronic mean free path becomes greater than the classical skin depth at very low temperature and high frequency. Comparison with Pippard's experiments at longer wavelengths gives fairly good, but not complete, checks of the theory, both for normal and superconductors.

Microwave Experiments with Superconductors

By A. B. PIPPARD

Royal Society Mond Laboratory

The surface resistance of normal and superconducting tin has been measured at a wavelength of 3.2 cm. Comparison with earlier measurements at 25 cm. shows that just above the superconducting transition temperature the resistance varies as $\omega^{2/3}$, in accordance with the theory of Reuter and Sondheimer. From dimensional arguments, based on the "two-fluid" concept of superconductivity, as exemplified by the theories of F. and H. London, Heisenberg, and Born and Cheng, relations are deduced for the expected behaviour, as the frequency is changed, of the superconducting resistance just below the transition temperature and at much lower temperatures. These relations are not confirmed by the experiments, and up to now no additional resistive mechanism has been found to explain the discrepancies. It is possible, therefore, that the phenomenon of superconductivity may not be explicable in terms of a two-fluid model. (*Not to be published.*)

Vapour Pressure of Helium 3

By J. DE BOER

Amsterdam

The differences in the properties of isotopic substances like He^3 and He^4 and like H_2 , DH , D_2 and DT are due to the strong influence of quantum effects on the properties of these light substances.

An exact theoretical treatment of these quantum effects meets great difficulties, resulting from the impossibility of treating the many particle problem; in the solid state the large amplitude of the zero-point vibration makes application of the usual theory of solids very doubtful, and in the liquid or condensed gas state, the known difficulties of treating fluid systems are added to the quantum effects.

The magnitude of these quantum effects in the condensed noble gases and the hydrogens can be determined in a semi-empirical way, avoiding the difficulties mentioned above by application of the quantum theory of the law of corresponding states. All physical quantities are reduced by means of "molecular units", which are defined in terms of the length σ and the energy ϵ , characterizing the intermolecular potential field, and the values of which are known for the sequence of substances: Xe , Kr , Ar , Ne , H_2 and He .

In classical theory the reduced equation of state should be the same for all these substances. Quantum mechanical considerations show, however, that the reduced equation of state should be a function of the "quantum parameter" $\Lambda^* = \hbar/\sigma\sqrt{m\epsilon}$ in which the mass m of the molecules occurs. Indeed, the reduced values of all physical quantities characteristic of condensed substances: molar volume and heat of vapourization, triple point—and critical temperature, Debye-temperature etc., appear to be smooth functions of this quantum parameter, the limiting value for $\Lambda^* \rightarrow 0$ being the classical value.

This dependence of reduced physical quantities on Λ^* can be used to calculate unknown values of these quantities for isotopic substances He^3 or HD , D_2 and DT , from the known experimental values of these quantities for the main isotope He^4 or H_2 respectively, because of the fact that molecules of isotopic substances have the same interaction. In this way the properties of liquid He^3 have been predicted, although the fact that in this case the functional dependence on Λ^* has to be extrapolated might lead to appreciable errors. The results are discussed in detail.

The critical temperature and pressure should be 3.5° to 3.1°K . and 1.3 to 0.95 atm . At 2°K . the vapour pressure of He^3 proves to be more than six times larger than that of He^4 . The vapour pressure formula is given and compared with that of He^4 . The results of the measurements of Fairbank *et al.* are discussed and prove to agree with the expected theoretical values. The deviations from Raoult's law will be discussed in detail. (*To be published in Physica.*)

The Separation of Helium 3 by Liquid Phase Methods

By J. G. DAUNT

Ohio

A general review is given of the experimental work carried out on the properties of the helium isotopes in the liquid phase. Detail is given of the experiments which established that the superfluid flow of matter in liquid helium II is one which does not involve the rare isotope He^3 and that in consequence a new method of isotopic separation is available.

An account is given of experiments which established the existence of an osmosis occurring between solutions of He^3 in He^4 of different concentrations and which have been carried out on the measurement of the distribution coefficient in such solutions.

An account is also given of the technique of separation now set up for continuous operation.

Experiments on Helium Films

By J. B. BROWN

Oxford

Experiments have been made to investigate the dependence of Helium film transport rate on the vapour pressure over the film. The film has been isolated from the bulk liquid and the vapour over it by a technique employing capillary leaks in wire packed tubes.

A study has also been made of anomalously high film transport rates through holes in surfaces. The effect has been related to previously observed high transport rates near the rim of cylindrical beakers and it is seen to have a practical bearing on the design of baffles used for attaining low temperatures by pumping the vapour over liquid helium.

LETTERS TO THE EDITOR

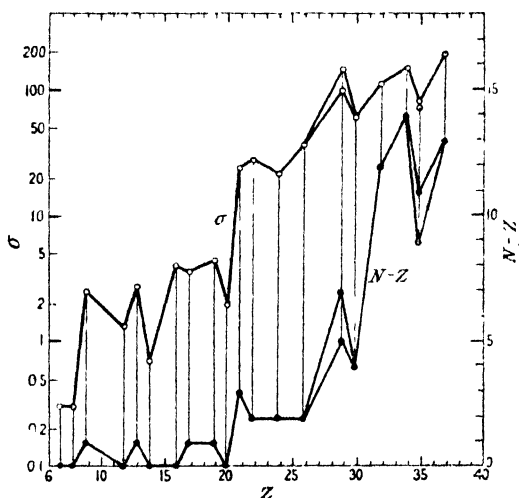
Nuclear Photo-Effect and Isotopic Number

The first measurements of the yields of neutron nuclear photo-effects (γ, n) by Bothe and Gentner have been refined and extended to a greater number of elements by Scherrer and his collaborators. Recently, Wäffler and Hirzel (1948) have published new values of the relative (γ, n) cross-section σ of 35 isotopes with atomic numbers Z reaching from 7 to 79 for the lithium γ -rays. In the figure these values of σ are represented up to $Z = 37$ (upper line, logarithmic scale).

The general increase of σ with increasing Z may be explained by the fact that the level density of nuclei for the high excitation energy in question increases rapidly with Z . In order to account for the remarkable rise of σ for $Z > 20$, Wäffler and Hirzel tentatively assume a corresponding rise of the level density in this region of Z .

Looking for further reasons for this sudden rise and for the fluctuations of σ for the light nuclei, I have compared σ with the isotopic number $N - Z$ of those isotopes from which the (γ, n) reaction starts (N = neutron number). The lower line in the figure represents $N - Z$ on a linear scale. A striking parallelism is seen which is only interrupted in the two cases of ^{32}S and ^{48}Ti , so that it can hardly be regarded as accidental. Beyond $Z = 40$ the correlation between the two quantities is no longer evident.

One might first attempt to explain this correlation by means of energy considerations, since light nuclei with $N - Z \geq 1$ have a comparatively low neutron binding energy and therefore the neutron emission probability of these nuclei for the excitation energy 17 Mev. is comparatively large. It should then be assumed that nuclei with a smaller isotopic number would emit a correspondingly greater number of protons and α -particles. Too little is known about the ratios of the emission probabilities of neutrons, protons and α -particles for the single isotopes of the light elements up to Ca, so that this assumption cannot as yet be checked directly. However, according to the measurements of the (γ, n)



Relative (γ, n) cross-section σ , according to Wäffler and Hirzel, and isotopic number $N - Z$, plotted against the atomic number Z .

and (γ, p) cross-sections by Wäffler and Hirzel (1947, 1948), it is known that for nuclei heavier than Ca the (γ, n) reaction is always by far the most probable one, independently of the isotopic number. Moreover, it is doubtful whether the (γ, p) process may be regarded as competing, according to Bohr's conception, with the (γ, n) process (Hirzel and Wäffler 1947, Schiff 1948, Jensen 1948).

Thus the simple consideration of binding energies is not sufficient to account for the correlation between σ and $N-Z$. One may then assume the nuclear structure to be also responsible. Accordingly, light nuclei should have a structure which makes them, other things being equal, more capable of absorbing the Li γ -rays and (or) emitting neutrons the greater their "neutron excess" $N-Z$. Here the domain of "light" nuclei should, according to the figure, be extended to about $Z=40$.

The author is indebted to Professor Gentner for valuable discussions.

Physical Institute of the University
of Freiburg, Germany.
September 1948.

P. JENSEN.

HIRZEL, O., and WÄFFLER, H., 1947, *Helv. phys. Acta*, **20**, 373.

JENSEN, P., 1948, *Naturwissenschaften* (in the press).

SCHIFF, L. I., 1948, *Phys. Rev.*, **73**, 1311.

WÄFFLER, H., and HIRZEL, O., 1948 *Helv. phys. Acta*, **21**, 200.

Validity of Two-Dimensional Design of Synchrotron Pole-Faces

In a recent note Davy (1948) has considered the design of a synchrotron pole-face using a two-dimensional theory. His result may be generalized by the statement that, to obtain a field $H_z : H_0(x/r_0)^{-n}$ on the $z=0$ plane, pole-faces are required which satisfy the equation

$$r = \text{constant} \times \text{cosec}^{1/1-n}[(1-n)\theta]. \quad \dots\dots (1)$$

x and z are cartesian coordinates as shown in figure 1 (b), while r and θ are the corresponding polar coordinates; H_0 is the value of H_z at $x=r_0$, r_0 being the equilibrium orbit radius. Equation (1) is simpler than the exact solution of the three-dimensional problem by Coggeshall and Muskat (1944), and, indeed, it is permissible to approximate still further by assuming θ to be small and to obtain for the pole-faces the equations

$$z = \pm \text{constant} \times x^n. \quad \dots\dots (2)$$

Equation (2) differs imperceptibly from equation (1) in practical designs and has been used as the starting point in the empirical design of pole-faces of the 30, 140 and 375 mev. synchrotrons being constructed in Britain (Fry *et al.* 1948).

To examine the validity of a two-dimensional design, the exact profile could be calculated, following Coggeshall and Muskat, and the errors of the simpler design assessed. The effect of fringing fields and of the exciting coils is important, however, and in a betatron-started synchrotron, the betatron-core affects the field between the main poles. It is therefore more realistic to compare two-dimensional measurements in an electrolytic tank (Peierls 1946) with measurements on a three-dimensional magnet. As an example, figure 1 (a), curve I, shows measurements of n , taken along $z=0$ in a two-dimensional electrolytic tank, on the pole-faces of the 30 mev. synchrotron. These measurements should be compared with the magnetic measurements shown by curve II. The differences between the two- and three-dimensional measurements are not very marked, and a sufficiently correct design can be made from two-dimensional electrolytic tank measurements. It is not possible, in any case, to obtain a constant value of n over the entire area between the pole-faces, as is shown by the "iso- n " lines of figure 1 (b); the important consideration is the region where stable orbits are possible, i.e. the regions for which $0 \leq n \leq 1$. It should be noted that the values of n in the figures are defined by

$$n = -(x/H_z)(\partial H_z / \partial x). \quad \dots\dots (3)$$

This is the most convenient definition in experimental work, although it differs from that used in equations (1) and (2) unless x/r_0 is near unity and θ is small.

The 30 mev. synchrotron pole-face gives perhaps the greatest divergence between two- and three-dimensional measurements which is likely to be met in practice; the pole-face

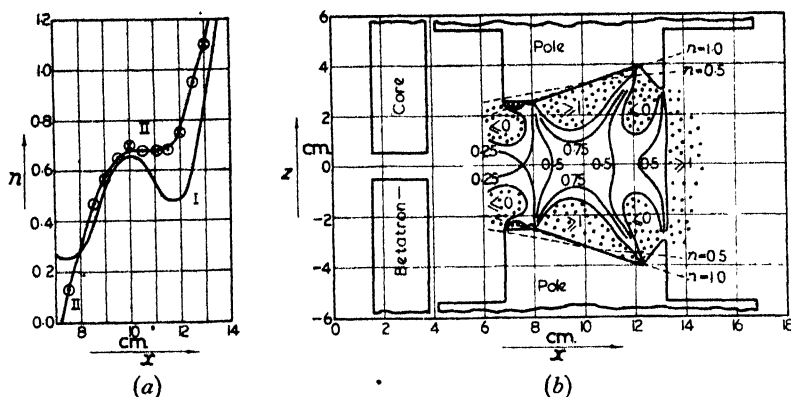


Figure 1. Comparison of electrolytic tank and magnetic measurements for 30 mev. synchrotron.

- (a) Curve I shows two-dimensional measurements in electrolytic tank. Curve II shows three-dimensional magnetic measurements, dots indicating D.C. measurements and crosses A.C. measurements.
- (b) Pole-face profile showing lines of constant n from electrolytic tank measurements. The broken lines show the pole-face contours given by equation (1). The dotted regions are those where the electron orbits are unstable.

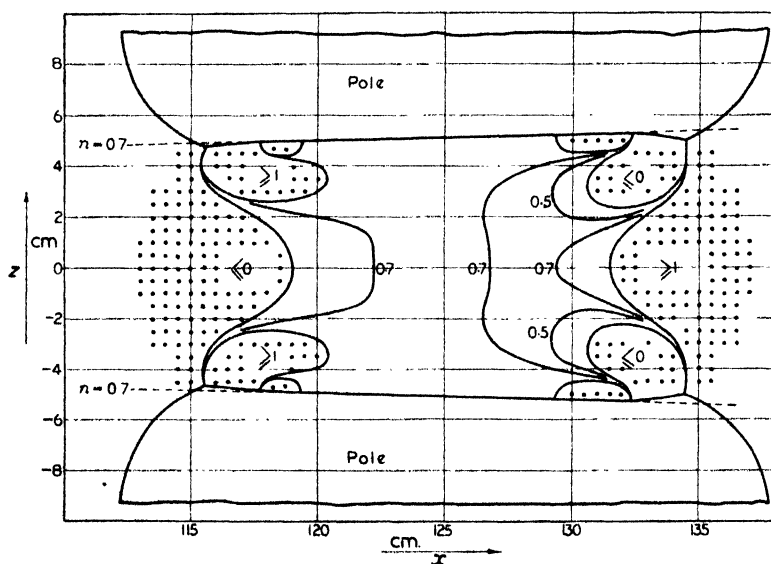


Figure 2. Electrolytic tank measurements for 375 mev. synchrotron.

extends over a radial distance of more than half the equilibrium orbit radius and θ takes values up to 20° . By comparison with the 30 mev. synchrotron the higher energy machines give a very small discrepancy between two- and three-dimensional measurements. Thus the 375 mev. synchrotron pole-face shown in figure 2, together with the "iso- n " contours,

was designed by a two-dimensional electrolytic tank, but it has not been possible to measure any deviations between magnetic measurements and the measurements of figure 2.

Figures 1 (b) and 2 also show the comparison of the solutions of equations (1) or (2) with the empirical designs. In figure 1 (b) the solutions for $n=0.5$ and $n=1.0$ are given by the dotted lines, and it may be seen that the lip correcting for fringing has an important effect on the slope required for the main part of the pole-face. In figure 2, however, θ is very much smaller and fringing effects are less pronounced; in this instance, the main pole-face contour cannot be distinguished from the solution of equations (1) or (2) for $n=0.7$, shown by the dotted lines.

We are indebted to J. Dain, of the Atomic Energy Research Establishment, and to J. F. Fowler and J. F. Smee, of the Metropolitan-Vickers Electrical Company, for permission to refer to magnetic measurements taken by them. The work described in this note was carried out in the Electronics Division of the Atomic Energy Research Establishment and is published by permission of the Director.

A.E.R.E. Division,
Telecommunications Research Establishment,
Malvern.

F. K. GOWARD.
J. J. WILKINS.

9th August 1948.

COGGESHALL, N. D., and MUSKAT, M., 1944, *Phys. Rev.*, **66**, 187.

DAVY, N., 1948, *Proc. Phys. Soc.*, **60**, 598.

FRY, D. W., GALLUP, J. W., GOWARD, F. K., and DAIN, J., 1948, *Nature, Lond.*, **161**, 504.

PEETERS, R. E., 1946, *Nature, Lond.*, **158**, 831.

REVIEWS OF BOOKS

Radio Aids to Navigation, by R. A. SMITH. Pp. xii+114. (Cambridge: University Press, 1947.) 9s.

During the last war striking advances took place in the application of radio techniques to navigation. For the most part these advances arose naturally through the growth of radar, and they included the various radar beacon systems, e.g. precision blind-bombing devices such as Oboc and G.H., the hyperbolic lattice systems such as Gee and Loran, and, later, the use of microwave radar to obtain maps of the country over which aircraft were flying. Much research leading to these developments took place in the laboratories of the Telecommunications Research Laboratory, and it is fitting therefore that Dr. R. A. Smith, himself a member of T.R.E., and an acknowledged authority in this field of knowledge, should have written this book.

The book records the advances made during the war years in both radar and continuous-wave methods of navigation and attempts to give a balanced survey of the whole field. The treatment is non-mathematical and no attempt is made to describe any of the systems in detail. Rather the author attempts to outline the various systems of radio (and radar) navigation developed during the last war, indicate their essential characteristics, and make critical inter-comparisons. Throughout, the author stresses the principles of the subject and comments on the systems both from the point of view of the technical design and from the point of view of the user. The book is well written and easy to read. The chapters include discussion of pre-war navigational systems, of the short and long range navigational aids—both radar and c.w.—developed during the last war, of microwave systems, of precision bombing aids, of radio altimeters, and of aids to instrument landing.

The book will be of interest to a wide range of readers, particularly all those who wish to learn something of the wartime developments in this field with a view to taking an intelligent interest in the applications in the field of civil flying and marine navigation. The book can be confidently recommended.

DENIS TAYLOR.

The Atom and its Energy, by E. N. DA C. ANDRADE. Pp. viii+196. (London: G. Bell and Sons Ltd., 1947.) 10s.

The purpose of this book is to give an account, for the reader not trained in the physical sciences, of those principles and discoveries in physics which have made it practicable to derive energy on a useful scale from the nucleus. The book, the first few chapters of which embody much of an earlier book by the same author, begins with the considerations that led to the atomic theory of matter. It then reviews the evidence on which Rutherford based his hypothesis of the nuclear atom and that on which modern ideas of atomic and nuclear structure depend. The last chapters are devoted to nuclear transmutation, fission, and the chain reaction and its consequences. There is a clear exposition of the principles and advantages of the proposals made by the Lilienthal Committee for the international control of atomic energy.

Professor Andrade is wholly successful in achieving the aim that he sets himself. His book is remarkable for its comprehensiveness and clarity, a clarity that owes much to the aptness of the author's down-to-earth analogies and illustrations, based as they are on notions readily conceived by the layman. The reader's feet are kept firmly on the ground by the author's custom of translating into everyday terms quantities that are often left in theoretical, or at least unfamiliar, units. These qualities of Professor Andrade's, and his robust yet scholarly prose, will be expected by those familiar with his writings. They combine to make this one of the best non-specialist books on the subject that the reviewer has seen.

The book is well made and printed, and there are comparatively few printer's errors. One or two minor slips, which might be misleading to the lay reader, should be mentioned so that they may be corrected in later editions. On page 72 the reference to "a smaller number whose mass was only 35 units" should read "a larger number . . .". On page 148 the statement that the lightest uranium isotope has mass 234 should be qualified by describing it as the lightest naturally occurring uranium isotope, otherwise the later reference to U233 may be confusing. On the same page, a few lines further on, it is said that there are 82 protons in the nucleus of Ba138; this should read 82 neutrons.

K. E. B. JAY.

Electrons (+ and -), Protons, Photons, Neutrons, Mesotrons and Cosmic Rays, by ROBERT A. MILLIKAN. Pp. x+642. (London and Chicago: The University of Chicago Press, 1947. Agent: Cambridge University Press.) Second Edition. 30s. net.

The present volume is a revised edition of the 1936 issue, in some respects brought up to date. It is by no means as comprehensive as the title makes one believe since many important aspects connected particularly with neutrons and protons are not even mentioned. The book might be used by a student for preparation of his final examination, but he would certainly fail if it was the only source of information. In spite of this it strikes me as a very valuable contribution, but ought to be called "Professor Millikan's Contributions to Modern Atomic Physics". With this title it is even more impressive, as it shows how much we owe to Millikan. The historical approach as seen from the Norman Bridge Laboratory of the California Institute of Technology in beautiful Pasadena is very instructive; the arguments which form the basis of the present ideas of atomic physics, too often taken for granted by the modern student of physics, are mostly presented in a dramatic manner. This and the complete absence of mathematics make the book very readable even for non-physicists (the appendix contains some derivations of formulae used in the text; the little space devoted to this only permits to make a result plausible, but certainly gives no hint of theoretical background). It is a little surprising to find that there are still a number of free positive electrons in the nucleus (appendix F, p. 620).

Picking out for special consideration chapter XVI on the release and utilization of nuclear energy, one can expect some strong comments from Professor Millikan. Section 1 on the packing fraction is devoted to the origin of solar energy production and of that of the cosmic rays, section 2 to radioactivity and fission. The real fun begins with section 3 on the control of the atomic bomb, a section which includes very pertinent considerations on the use of atomic power. Millikan does not believe that atomic energy can ever become a major source of energy because the processing cost of nuclear fuel is very high, whilst

coal is ready for use. Moreover, he makes the point that to use "atomic packing fraction energy" is very objectionable as it involves using up rare and valuable elements such as U and Th. He says: "In any case, intelligence demands the conserving, rather than burning up, of a very rare and precious resource". As far as the atomic bomb is concerned, Millikan says: "The great service to mankind of the advent of the atomic bomb has been to make as clear as crystal, to all classes and conditions of men the world over, the necessity to put a stop to international war".

Chapter XX contains, in some 60 pages entitled the Atom-Annihilation Hypothesis as to the Origin of the Cosmic Rays, much of Millikan's cosmic ray experiments and speculations in a very persuasive form.

It would be easy to throw stones and even heavier objects at the author when one goes into a discussion of historical justice. I am not qualified for this, and at any rate would consider it very unfruitful to enter into the jungle of controversy. E. BRETSCHER.

Cosmic Rays and Nuclear Physics, by L. JÁNOSSY. Pp. xiv+186. (The Pilot Press, 1948.) 9s. 6d.

The "Frontiers of Science" Series, published by the Pilot Press, in which Professor Jánosy's book appears, aims "to set a new standard in popular science . . . intended for the growing non-specialist public which wants to keep abreast of recent scientific developments . . .". To do justice to *Cosmic Rays and Nuclear Physics* this body of readers must be identified from the internal evidence of the book, and the term "non-specialist" must be accepted in a rather rigorous sense. This book may be recommended strongly to all who have recently taken an honours degree in physics; it will be of the greatest value to students now reading courses which include nuclear physics (for as the title aptly suggests, the cosmic-ray physics of today promises to be the father of the nuclear physics of tomorrow), and, indeed, would be essential to all research workers in the cosmic-ray field were it not for the recent publication of Professor Jánosy's more exhaustive work. Those with less specialist equipment are likely to find difficult reading, involving a complexity of theme and demanding a mature if elementary mathematical sense. Their task is made the more difficult by a number of weaknesses in the diagrams and in the setting of the mathematical matter which, while they do not seriously trouble the reader who is familiar with the subject, are likely to be a serious hindrance to others.

The treatment is, on the whole, conservative and conveys admirably the form of argument involved and some of the pitfalls encountered in the cosmic-ray physics of the last decade. The chapters on the Analysis of Cosmic Ray Particles, on the Theory of the Meson and its Instability, and on the Geomagnetic Effects will be particularly stimulating to the serious reader, and all cover excellently matter which has not been well described at this level. The treatment of the Soft Component is less successful—a more physical and less mathematical approach would probably have been more satisfactory—while the chapter on Penetrating Showers is detailed to an extent which the topical interest of these events in cosmic-ray laboratories at the moment cannot completely justify. The treatment throughout, however, has the authority which can only come from close contact with the crucial problems of the subject as they have arisen, and the balance of the discussions of experimental work is very sound.

The illustrations are an unsatisfactory feature of the book. Too many line diagrams are insufficiently described, while a large proportion of the half-tone illustrations are too small to be of value to anybody not familiar with the originals. In particular, the line diagram figure 41 has been allowed to appear over a most unfortunate title. The classical apparatus of C. T. R. Wilson, the functional excellence of which has always been apparent to later cloud-chamber workers, was not remotely like this illustration, which should be replaced in any later edition.

J. G. WILSON.

Obituary Notices

ALEXANDER MÜLLER

DR. ALEXANDER MÜLLER, who died on 2 July 1947 at the age of 58, was well known for his work on x rays. He began his studies in his home town of Zurich, where he specialized in mathematics and physics and was awarded a doctorate for a thesis on "Messungen der thermischen Ausdehnung von kristallisierten Quartz und von Gold zwischen 18° und 540° ". Later he worked with Debye in Germany and in 1919 came to Cambridge to work under Lord Rutherford.

By this time his main interest was in x rays. He had already made one of the earliest determinations of the value of Planck's constant, using an x-ray method, and had become interested in x-ray spectroscopy. At Cambridge he continued with his work on x-ray spectroscopy and published a paper on the L spectrum of mercury for which work he designed and made a special x-ray tube with a liquid mercury anticathode, and a photographic spectrometer which, while capable of high accuracy, was a model of compactness. Later he was consulted by Professor G. I. Taylor, who was interested in the slip phenomena of single metal crystals, and he developed a method for the study of such crystals by x-ray diffraction; the results obtained in this way formed the basis of the Bakerian Lecture by Taylor and Elam in 1923. In the course of this work he himself made many sodium and lithium single crystals and experimented on their deformation.

His colleagues at the Cavendish Laboratory were quick to realize that in Müller they had in their midst an experimentalist of outstanding ability and one who combined a sound theoretical knowledge with an acute critical faculty. Although his work was rather apart from the main radioactive researches of the laboratory he was continually being consulted by other workers on their own particular problems.

In 1922 he left Cambridge to join Sir William Bragg's research staff at University College, London, where Bragg was collecting around him a research group whose main aim was the application of x-ray methods in the study of crystal structures. About this time Bragg had just broken new ground by his pioneer work on the structure of naphthalene and anthracene, the first organic structures to be investigated by x-ray methods. Up till then Bragg and his colleagues had used for the most part the Bragg ionization x-ray spectrometer. Müller brought to the work his extensive knowledge of photographic methods, and it was not long before such methods were being used to supplement the ionization methods which still had many advantages, especially in the study of single crystals. For powders and crystal aggregates the photographic technique was both an easier and a more reliable method of investigation.

In 1923, Bragg was appointed Director of the Royal Institution and took with him from University College a team of workers to provide a nucleus of a Research school in the Davy Faraday Laboratory attached to the Royal Institution. Müller was one of this team and it was destined that the rest of his life's work should be carried out there.

The writer can well remember his first visit with Sir William Bragg to the Laboratory. It was very nearly deserted; in most rooms there was a collection of discarded and broken apparatus, accumulations of stones, presumably the object of study of some bygone mineralogist, and similar débris. There was no equipment of the type required and the workshop facilities were of the poorest. This was soon changed, and in the fitting out of the Laboratories and workshop Müller played an important part. He threw himself with his infectious enthusiasm and zeal into the task, and before long a collection of workmanlike laboratories was available and a new workshop was turning out first-class spectrometers and other equipment. Even in those early days Müller arranged for a vacuum service to be laid on to the various laboratories.

There was none too much money available for research in those days and at any rate there was little of the specialized x-ray equipment required that could be bought in the

open market. Müller's contribution in overcoming these difficulties was a notable one. He made his own high tension coils, interruptors, high vacuum pumps, and x-ray tubes, and taught others to be equally self-reliant, although few of his colleagues ever got quite as high efficiency from their apparatus as did Müller. Still, with Müller at hand to help, and he was always ready and willing to do so, they produced results.

His best known researches at the Davy Faraday Laboratory were connected with the structure of long-chain organic compounds. Just as Bragg had opened up a new field in his study of the aromatic hydrocarbons, so Müller, at first in collaboration with Shearer and Saville and later on his own, broke new ground by his studies of the aliphatic compounds. His work on the normal fatty acids and hydrocarbons (published in the *Journal of the Chemical Society*, 1923-25) established for the first time the dimensions and details of structure of such molecules, and this work was elaborated later when Müller succeeded in growing single crystals of some of these compounds, which enabled details of structure to be determined which could not be elucidated by studies of the highly oriented powders used in the earlier investigations. His later researches on these materials, including work on the Van der Waals forces, lattice energy, dielectric polarization and compressibility, were published in a series of papers presented to the Royal Society between 1928 and 1941. At the same time he was taking an active part in directing and supervising the researches of the large group of research workers in the Davy Faraday Laboratory of which in 1928 he had been made Assistant Director.

Another important contribution made by Müller during these years was the development of high power x-ray tubes. It was becoming increasingly important to be able to obtain x-ray diffraction photographs of substances in short times, especially for the study of thermal changes, elastic properties and the like. Müller designed and had made the first really efficient high power x-ray tube. This could deal with a continuous input of 5 kw. and was of the continuously evacuated high vacuum type with a water-cooled rotating anticathode. There were many difficulties in the construction of such a tube, mainly concerned with the maintenance of a high vacuum and the removal of the heat generated by the discharge, but these Müller with his customary ingenuity overcame. Tubes of this type were in continuous operation at the Davy Faraday Laboratory for many years, being used, for example, in connection with researches on the thermal vibration of atoms in crystals and on elastic properties of hydrocarbons. Later, in conjunction with Clay, he embarked on a much more ambitious project, the construction of a 50 kw. tube. This was an engineering proposition and involved not only the design and construction of the tube with its huge rotating anticathode some two feet in diameter, but also a three-phase generator using continuously evacuated rectifying valves. This task was successfully completed and an account of the outfit was published in the *Journal of the Institution of Electrical Engineers* in 1939.

During more recent years Müller's health left much to be desired, but in spite of this he carried on cheerfully with his work and was involved in various researches connected with problems arising out of the war.

Müller will be remembered with affection by a large number of workers who passed through the Davy Faraday Laboratory for the kindly way in which he assisted in their researches and for the valuable advice always so willingly and efficiently given and so often enlivened by a touch of humour. Not only was he a very sound physicist and a more than adequate mathematician, but he had wide scientific interests. It is probable indeed that had he been given the option of choosing his career again he might well have decided to study the more physical aspects of biological science. He had many other gifts. He was a first class linguist; English, French, German or Italian he spoke with nearly equal fluency. He was interested in art and music, could perform adequately on many instruments and was a good singer with an amazingly wide repertoire of songs of many countries. Many who made the trip across Canada with the British Association in 1927 will remember the impromptu concerts on the train and Müller's performances, instrumental and vocal. He was a keen mountaineer and spent most of his holidays in his native mountains, although he had also in his heart a warm spot for the hills of Skye.

He was elected a Fellow of the Physical Society in 1925.

SIR CLIFFORD COPLAND PATERSON, O.B.E., D.Sc., F.R.S.

THE death of Clifford Copland Paterson at the age of 68 removes a personality from our midst who has had a notable influence on the application of the developments of physical science to technological purposes.

After being at Mill Hill School in the time of the Headmastership of Sir John McClure, he was at the Finsbury Technical College and at Faraday House when it was under the guidance of Alexander Russell, of whom he always spoke with admiration and affection. He worked for some time with George Wailes & Co., General Engineers, and he supplemented his training at Faraday House in the works of the Maschinen Fabrik Oerlikon, near Zürich, which gave him a valuable experience of the practical side of electrical engineering. The leading Swiss engineering firms were among the most capable in Europe, no doubt due, in part, to the training available at the Technische Hochschule at Zürich, which young British engineers used to attend a generation previously, as there was no institution of similar standing in this country. During this period he worked for the firm in erecting the electric equipment in the tunnel of the Jungfrau Railway, an experience which gave him opportunities for walking and climbing in Switzerland; this meant to him, as to many others, a renewal of health of body and spirit which few experiences can equal. This varied training in general engineering served him in good stead in his later life in giving him an intuitive basis for the guidance of himself and others as to how the practical requirements of a problem might best be tackled.

Soon after the foundation of the National Physical Laboratory he was chosen by Dr. R. T. Glazebrook, the Director, to undertake the technical side of electrical work. The more fundamental researches, including the determination of the values of resistances and standard cells in terms of the "absolute" c.g.s. units, and of inductance and capacitance, were in the hands of F. E. Smith and A. Campbell. The most important part of Paterson's early work was the development of equipment for determining the accuracy of commercial types of apparatus submitted to the Laboratory for verification, including voltmeters, wattmeters and ammeters for alternating current. The present writer, after engineering experience in Manchester and Rugby, joined the Laboratory Staff soon after Paterson's appointment, and the design and installation of what has proved to be the most accurate and flexible system in any national laboratory for the measurement of alternating current quantities was one of the results of several years of close collaboration. The essential instruments, a voltmeter and a wattmeter, both of the electrostatic type, were developed so as to provide an accuracy of a few parts in 10,000. In those early days there was no public electric supply available at the Laboratory, and one remembers having to test three-phase instruments when the only power supply available was an overloaded gas engine.

Another branch of the electrotechnical work was the development of accurate measurements of candle power of light sources and the design and installation of standards of adequate luminous stability. The "Vernon Harcourt" Pentane Lamp was one of the recognized standard lamps in use, especially in the gas industry, and as it had also a nominal value of about 10 candle power, which was a more convenient size than working standards of about 1 candle power, it was adopted as a reference standard. As it was a flame lamp, the illumination from it was dependent on atmospheric conditions, particularly barometric pressure and humidity, and the influence of these had to be accurately determined. Moreover, it was found that lamps varied appreciably in the amount of light which they emitted, partly due to small variations in methods of construction. This had been only partly recognized previously, as the recently introduced contrast photometers were capable of detecting smaller differences between light sources than earlier types of instrument. In those days the general source of domestic illumination was the carbon-filament lamp, requiring about 4 watts per candle, compared with modern lamps of 1 watt and less. B. P. Dudding, J. W. T. Walsh and W. A. Barnett were closely associated with Paterson in this work, the development of which has continued to the present day, with the requirements imposed by the evolution of more efficient and "whiter" sources of light, involving the comparison of light of different colours.

In later years Paterson and the Staff of the Laboratories of the General Electric Co., gave most valuable and willing assistance to the Staff of the N.P.L., providing lamps specially designed as photometric standards, the construction of which required the technique and manufacturing methods only available in a lamp factory.

This led to the accurate intercomparison of the relative size of the unit of light in use in other countries—America, France and Germany—by circulating lamps among their national laboratories. The next step was to adopt as the recognized unit of candle power such a value that a lamp measured in one country would be assigned the same value when tested in another ; this was largely brought about by the efforts of the International Illumination Commission, of which Paterson was one of the most influential members.

This was followed by the decision to develop an International Standard which should not be dependent upon electric lamps which might change slowly or be destroyed ; and during the inter-war years it was left to the national laboratories of England and America to collaborate on a more fundamental unit dependent upon the brightness under “ black-body ” conditions of a surface at the melting point of platinum. The necessity for high accuracy made this a difficult piece of work, and success in this country was largely due to H. Buckley.

In all this work, some of it of a semi-political character, involving in some countries the supersession of long established standards, Paterson’s support was very helpful. The international recognition of this work was delayed for several years by the war, and it was only on 1st January 1948 that the “ New Unit ” came into official use at the same time as the “ absolute ” C.G.S. units of electrical quantities which superseded the previous empirical “ International ” Units, forty years after the change was urged by this country and America at the International Congress of 1908, but which did not receive unanimous support.

The work of Paterson and others at the National Physical Laboratory established the science and art of photometry on a sound basis in this country. By discussing the various problems involved with instrument makers and lamp manufacturers difficulties were gradually overcome.

During the war, 1914-1918, the results of development work in other countries, particularly Germany, were no longer available. It was a period of rapid change in the material used in lamps, carbon giving way to tantalum and tantalum to tungsten.

It was obvious that more research and development work was required in this country ; and Paterson’s advice as to how this should be initiated was sought by the Directors of the General Electric Co. As a result of the discussions of the organization which he suggested, he was invited to undertake the establishment of a Research Organization which he eventually accepted, resigning from the N.P.L. at the end of 1918 and undertaking his new duties on 1st January 1919.

National work of some importance which arose shortly before Paterson retired from the N.P.L. to join the General Electric Company was in connection with a committee appointed by the Board of Trade to advise on Navigation Lights of Merchant Ships at sea. International general rules for navigation lights had been in operation for some time, but no detailed technical specification of performance of the various types of lamps had been produced. The principal requirements were to determine the effective candle power required to enable an oil lamp to be seen at a distance of two miles in the case of the coloured starboard and port lights, and at five miles in the case of the mast-head lamps. A lens system had to be designed to enable this to be attained, providing adequate vertical divergence of the beam to enable a ship’s light to be seen when heeled over several degrees, a factor of great importance in the case of sailing ships. There were also the colour intensities of the red and green lights to be specified. The writer was a member of the committee and carried out with Messrs. Walsh and Buckley most of the investigatory work involved. Paterson became Chairman of the Committee in its later stages and showed in these comparatively early days that, by his clearness of exposition and friendly appreciation of the views of others, sometimes without much foundation of a scientific character, he could persuade them without difficulty to agree with his proposals when he considered them important.

He was Joint Secretary with D. R. Wilson (later Sir Duncan Wilson) of the Home Office Committee on the Lighting of Factories and Workshops which made its Report in 1915 ; and he was later Chairman of the Committee of the Building Research Board on the Lighting of Buildings.

An influential side of Paterson’s activities was his association with the professional and technical organizations, British and foreign. He and the writer carried out work in the early years of the N.P.L. on behalf of the Engineering Standards Committee which later developed into the British Standards Institution. One of his early tasks in this direction was the investigatory work on the performance of commercial electric lamps, the results of

which formed the foundation of the original specification of performance of (carbon) filament lamps. Since those days, up to the present year, we have been continuously associated with the work of the Electrical Industry Committee of the Institution and of its various Sub-Committees. Paterson was Chairman of this Committee for some years until he was elected Chairman of the Engineering Divisional Council and later Chairman of the General Council of the Institution. His election to such important positions was an indication of the appreciation of his friendly advice and of his understanding of different points of view as well as of his knowledge of engineering. His services in maintaining the quality of the work of the B.S.I. have been put on public record by the Director, Mr. Percy Good.

An indication of his activities and of the esteem in which he was held is shown by the fact that he had been President of the Institute of Physics, the Illuminating Engineering Society, the Junior Institution of Engineers, the Institution of Engineers-in-Charge and the Electrical Research Association. He was elected President of the Institution of Electrical Engineers for the year 1930-31, during which the Centenary of Faraday's discovery of electromagnetic induction, the foundation on which all the technical applications of electricity have been developed, was celebrated on an international scale. No happier choice could have been made.

On this, and many other occasions, he showed that he was an unusually effective lecturer, particularly to audiences of a popular character. He would describe physical phenomena in simple language, illustrate them by analogies which could be easily understood and proceed to the experiment which illustrated the phenomenon under discussion. He might refer to the transmission of speech by using radio waves or waves of light as the carrying agent, or the history of various forms of sources of light, or one of many other applications. The audience was sure to have both a widening of their knowledge and an intellectual treat of absorbing interest.

Reading through the scripts which he used at his most recent lectures one is struck by the marks in nearly every line indicating the varying degree of emphasis to be used in order to render it as intelligible as possible to the audience.

Among the many papers and addresses which he gave, those contributed, some of them jointly, to the Physical Society, were: 1909, International Standard of Light; 1912, Point Sources of Light; 1915, Temperature and Colour Identity; 1915, Unit of Candle Power in White Light; 1932, Addresses on the use of Photo-cells; 1937, Guthrie Lecture, Appraisal of Lighting. He was a member of the Council of the Society from 1914-1919 and Treasurer from 1938-1946.

Instances of his more recent contributions of an influential character are an Address to the Parliamentary and Scientific Committee, a paper contributed to the Stockholm Congress on Management, the lecture at the Central Hall on "The Electron Liberated", on the occasion of the fiftieth anniversary of the discovery of the electron by J. J. Thomson, and a lecture to the Royal Society on the "Research Laboratories of the General Electric Co." These were given in 1947, shortly before his departure with Lady Paterson for some months to Australia, where he was assured a cordial welcome as the founder of a Research Laboratory of world-wide reputation, as an authoritative exponent of British schemes of standardization which have extended their scope far beyond what was dealt with even a dozen years ago, and also as a Director of an English engineering firm, employing some 50,000 people, which has close contacts with all parts of the Commonwealth and with many other countries. Not least was it an opportunity for many to meet one whose staff had been largely instrumental in all the developments, production and utilization of equipment required for various forms of radiolocation.

The Twenty-First Anniversary of the founding of the Research Laboratories was celebrated in February 1944 with such festivities as war permitted. Among the events was a lecture by Paterson on the "Confidential History of the Research Laboratories" in which he described their foundation, growth and activities. The lecture provided a fascinating account of organization and design and of what he considered were the proper relations which should be established between the Laboratory and the various departments of the Company which it was to serve.

The present writer, as one of his oldest friends, had the pleasant task of proposing the thanks of the meeting and in doing so informed the audience that an abiding memory of Paterson of thirty years ago was that he always carried a centre-punch in his pocket. To everyone's surprise Paterson stood up and produced a centre-punch, showing that though, as he sometimes lamented, he had no time to do practical work himself, yet an elementary engineering tool did at least provide a tangible memento of his earlier joys in handicraft.

Two events occurred during recent years which enabled the Staff of the Laboratories to show their friendly affection for him and their appreciation of his services. One was his election as a Fellow of the Royal Society. In an address to the Staff on that occasion he said how glad he was that this had happened, as he regarded it primarily as a recognition of the value of their work over many years in widening the knowledge of science and developing its applications. The other was the knighthood which had been bestowed on him. He was in doubt as to whether he should accept the honour and, when his intimate friends persuaded him how disappointed the Staff would be if they learnt that he had declined, he decided to accept. The reception of the news at Wembley can have left him in no doubt as to the rightness of his decision.

In his relations with his Staff he regarded himself as a fellow traveller in life's journey, always giving service as he expected it of them. He "clocked in" at the beginning of the day's work, as was the duty of all the rest of the staff, and in his daily contacts with all and sundry he showed that he regarded them as colleagues in the important labours on which they were engaged.

Paterson was ready to take bold decisions in the national interest. At one stage in the war he and his Senior Staff realized that large numbers of novel types of valves would very likely be suddenly required. It was realized that the usual routine procedure for obtaining orders, particularly of a speculative character, would cause too serious delays; and production groups were set up in the Laboratories, employing eventually some 700 women, who were taught the various technical operations of glass blowing, making glass-to-metal seals, cathode formation and many other arts, including the production of a very high vacuum and sealing-off.

Eventually the day came when operational and supply officers did require these valves very, very urgently, and did not know how to get them. Paterson disclosed his private store of trained skill and valves, and the relief of the Services can be imagined. He had saved six to eight months of training and development which would otherwise have been required for the provision of valves which were essential for the national safety at a most critical period of the war, more essential indeed than anyone originally imagined would be the case.

In the course of his address to the Staff in February 1944, already mentioned, Paterson gave a short outline of the development during the war, and in conclusion he said:

"Wherever one opens the book of our war activities there one finds a fascinating story of apparently impossible things accomplished and difficulties surmounted by the patience and determination of individuals drawn from all parts of the Laboratories and Shops, backed by girls and women in the Offices and Laboratories and working in Groups unselfishly for the one object. It is true that this has always been the tradition of the place from its commencement; but the 1000 newcomers to the Laboratories since 1939 had not had the years of apprenticeship of the older members nor, in pre-war times, have people had the appalling irritations and constant discouragements arising from Government controls, blitzes and other man-made hindrances to rational existence. Therefore I return to where I started in this talk, and thank Providence which has permitted me to share the comradeship of such a staff through the days of peace and still more wonderful days of war. I trust it is not too much to hope that I may still tread with you the adventurous paths of reconstruction, whenever our war task shall draw to a close."

It was hoped by his friends that the new experience of his journey to Australia, which commenced in the early days of 1948, would provide him with a well-earned holiday, in spite of a full programme which had been prepared for him, and that he would return with renewed zest to "tread with them the adventurous paths of reconstruction". However, it was not to be. He became seriously ill in Australia, surviving the return journey only a few days, and died on 26th July.

He was awarded the O.B.E. for his work with Walsh in developing an electrically indicating aircraft height finder in 1916-1917, and was made an Honorary Doctor of Science of Birmingham University in 1937. In 1946 he was given the James Alfred Ewing Medal, which is awarded jointly by the Royal Society and the Institution of Civil Engineers, and in 1948 he was the first person who was not an American to be awarded the Gold Medal of the Illuminating Engineering Society of America. In the summer of 1948 he was elected as Honorary President of the International Illumination Commission, which has distinguished no other person in this manner.

It is hardly possible, in giving an account of Paterson's life work, not to include Lady Paterson in one's thoughts and esteem. It was characteristic of her to insist, with her daughter, in joining the Laboratory Staff as a volunteer and beginner in valve construction and assembly when urgency for output was becoming very serious. She carried on this work for many months.

Her presence and kindly interest in the members of the staff was a continual source of encouragement and sympathy, and must have been a reassuring influence when many were engaged on delicate manipulative work, often interrupted by air raid warnings. At one period these warnings were to be expected daily, and the A.R.P. record shows that on one day ten were received and circulated to the staff.

Lady Paterson took charge of the War Savings Organization of the Research Laboratories, the funds of which have amounted to £110,000.

The mutual affection between Sir Clifford and Lady Paterson is hardly a subject which one can discuss, except to say that it was a living witness of two characters which can be an inspiration to all.

E. H. R.

DR. ALLAN FERGUSON writes: May I be permitted to add an account of two incidents which seem to me very characteristic of our old friend?

The first deals with an evening address which Paterson gave at a York meeting of the British Association in 1932. The address had drawn to its close and a spontaneous burst of applause was dying away when the enthusiastic voice of the President of the Association, Sir Alfred Ewing, was heard to ring out: "Oh, you incomparable showman." And without the slightest *arrière pensée* the President's almost instinctive shout of delight was a testimonial to Paterson's supreme craftsmanship. Yes, indeed, he was a great artist, with the art that conceals art. Every experiment marched to its appointed close with the utmost precision, and Paterson had no need to look at the lantern screen in order to see what was thereon. Paterson's task in any address which he gave was to make sure that the facts and principles should be brought out in the most telling fashion, and he never failed in this task.

The second incident, which illustrated another side of Paterson's character, occurred only a few months ago. We both arrived early for a Committee meeting and were talking of the world situation when I remarked that I could see no light anywhere. "Well", said Paterson, "there is at least a possibility of light". "In what?" "In a return to Christian ethics and Christian principles."

EMERITUS PROFESSOR S. W. J. SMITH, M.A., D.Sc., F.INST.P., F.R.S.

SAMUEL WALTER JOHNSON SMITH was born at Ferry-Port on Craig, Fife, in 1871, and was the second son of Walter Mackerzie Smith and of Margaret Black. His father was a pioneer in railway development in Japan and was responsible for many improvements in locomotive design, notably in connection with the compound steam engine. He was educated at Tayport Public School, West End Academy, Dundee, and Rutherford College, Newcastle-on-Tyne. From Rutherford College he won a National Scholarship to the Royal College of Science in 1887 which he held until he went up to Trinity College, Cambridge, with a Major Scholarship, in 1891. He obtained a first class in Part I of the natural science Tripos in 1892 and a first class in Part II in 1894, in both Physics and Chemistry. A first class in the two subjects is an unusual achievement and this equipped him for the two distinct branches of research which he later developed.

He became Coutts Trotter Student and began work in the Cavendish Laboratory under J. J. Thomson, choosing as his subject electrocapillary phenomena and the capillary electrometer. Smith, who was discriminating in his praise, always spoke with the greatest respect and admiration of Thomson and his work. This early interest in electrolytic

effects led to one of his major fields of investigation. He published a paper on "The nature of electrocapillary phenomena" in the *Philosophical Transactions of the Royal Society* in 1899; and a paper on "A portable capillary electrometer" in 1903, the first of a long series in the *Proceedings of the Physical Society*.

Smith had returned to the Royal College of Science as Senior Demonstrator in Physics in 1897. Sir Arthur Rücker, who was then the Professor of Physics there, suggested to him that an investigation of magnetism in meteoric iron might shed light on the nature of the earth's magnetism. This suggestion led to Smith's first work on magnetism which developed into his deep interest in the magnetic properties of materials, an interest which he retained throughout his life. An account of his early experiments is given in the *Philosophical Transactions of the Royal Society* of 1908 in a paper entitled "The thermomagnetic analyses of meteoric and artificial nickel iron alloys". He at the same time continued his work on electrolytic phenomena which was published in a number of papers in the *Proceedings of the Physical Society* of about that period. H. Moss and W. F. Higgins collaborated with him in parts of this work.

In 1912 Smith was made an Assistant Professor in what had now become the Imperial College of Science and Technology. This was a period of rapid and most fruitful development of his two main lines of investigation. His interest in magnetism and its dependence on temperature had now extended to the carbon steels and their constituents. Papers in the *Proceedings of the Physical Society* followed each other in rapid succession; "The magnetic transition temperature of cementite" (with W. White and S. G. Barker) 1911; "The self-demagnetization of annealed steel rods" (with J. Guild) 1912; "The thermomagnetic study of the eutectoid transition point of carbon steels" (with J. Guild) in the *Philosophical Transactions of the Royal Society* of 1914. In that year, 1914, he was made a Fellow of the Royal Society. His interest in ionic problems continued and in 1916 he published three papers dealing with this subject in the *Proceedings of the Physical Society*.

The conditions under which Smith was working at the Imperial College were particularly congenial to him; he was in charge of the advanced teaching laboratory and enjoyed talking to the relatively small and keen class of about twenty students. He was free from departmental administrative duties, and had ample time for his research work, with facilities and space to carry it out. His wholehearted devotion to his subject could be indulged to the full. He was on the most friendly terms with H. L. Callendar, who was now professor, and his family. This happy atmosphere made a lasting impression on him. In Smith's own words, "Callendar's friendliness and entire freedom from affectation of any kind, with no hint of superiority in anything he said or did, made a deep impression on everyone who knew him".

In 1919 Smith was appointed to the Chair of Physics at Birmingham. One of his first duties was to organize the teaching of a full honours school in the subject. This meant the introduction of new advanced courses in which he took a very large share. In his first few years at Birmingham he also undertook a good deal of the more elementary lecturing, particularly in electricity and magnetism, including the large first-year class. These first few years coincided with the return of large numbers of ex-service men and women from the first world war. This necessitated an extension to his department by the addition of laboratories housed in wooden huts. In all these developments he took a leading and active part.

As a teacher he was careful and exact. He enjoyed the development of a physical argument and to do this he always prepared detailed lecture notes which he wrote and re-wrote until the exact wording was to his liking. This meant that to him lecturing was no easy matter, and his method was probably most appreciated by the more advanced students to whom he lectured on thermodynamics and crystal optics, with also a complete course in the more recent advances in physics. His attitude to the announcement of new discoveries was cautious; he did not wish to appear fashionable or at the head of the field in such matters. He was, however, a great deal more interested and knowledgeable in the progress of new ideas than he was generally prepared to admit, and anyone would have been foolish to assume his ignorance of such things. When new ideas had been satisfactorily established they quietly appeared without delay in his own teaching.

Although as head of the large physics department at Birmingham he had a good deal of teaching and administrative work to do, his main interest still lay in his research. He now concentrated on magnetism and on the other properties of metals with which magnetism

is associated. In this connection he recognized the power of the new techniques provided by micro-photography and by x-ray crystal analysis and work was begun on both these aspects of his problem. In this he had the devoted assistance of two of his colleagues, A. A. Dee and the late James Young, who threw themselves energetically into the work. Dee concentrated on the thermo-magnetic analysis and the micro-photography, while Young developed the work with x rays. This extension of the research to the crystallographic side remained one of Smith's main interests, the problems of solid geometry which were involved being a source of delight to him. In 1924 he published in collaboration with A. A. Dee and M. V. Mayneord "The magnetism of annealed carbon steels" in the *Proceedings of the Physical Society*. In 1928 an important paper in three separate parts on "The mode of formation of Neumann bands" dealing with the later developments of the work, was published in collaboration with Dee and Young in the *Proceedings of the Royal Society*.

In addition to his normal duties as a professor, Smith voluntarily undertook while at Birmingham a number of additional responsibilities in connection with his department. These included an active part in government investigations on radio direction finding (under the Radio Research Board of the Department of Scientific and Industrial Research), and on the transmission over long distances of the sound of gun-fire with Dr. Guy Barlow (under the Meteorological Office). At the request of Sir Gilbert Barling, the Pro-Chancellor, on behalf of the National Radium Commission, he established a laboratory to provide radon for medical use throughout the Midlands. In all such matters he took a deep interest. He held the office of Dean of the Faculty of Science from 1929 to 1932 with dignity and distinction. Although he did not claim to be, or wish to be, a man of affairs, he brought to the office a fund of shrewd judgment and wide learning, and an extensive knowledge of University matters.

Smith retired from the Chair of Physics at Birmingham in 1936 and was made an Emeritus Professor of the University. He continued his interest in the magnetic and crystal problems, his last published work being a note, with Young, to *Nature* in 1929. "The Widmanstätten structure of octahedral meteoric iron". The subject on which Smith was working had become a highly complex one and it was no easy matter to correlate and explain the large mass of experimental data which he and his collaborators had collected. In his later years failing health prevented his completing this work and it was a matter of regret to him that the results of the work of a number of those who had helped him should not have been published.

Smith always took a special interest in the work of the Physical Society and held the office of Secretary from 1908 to 1916, and that of Vice-President from 1916 to 1919 when he left London for Birmingham. In later years his affection for the Society continued. One of the first meetings of the Society outside London was held in Birmingham in 1929 when a number of papers were read on a wide variety of subjects dealing with his own work and with other investigations which he had freely encouraged in his department. This was a very happy occasion for him.

Many who knew him will remember him for his friendliness and kindness. He was no respecter of personages but enjoyed talking to simple folk and especially to children. In these talks he would occasionally embarrass his young friends by asking questions on science or arithmetic. Smith liked to tell the following little story against himself. He had been asking a very youthful porter on a suburban station a number of such questions, when the boy neatly and politely extricated himself by saying "I think your train is coming now, sir".

He died on the 20th August 1948. In 1900 he married Dorothy Muriel Stamp. His widow and his daughter Jean survive him.

T. L. IBBS.

INDEX

	PAGE
Accelerators, linear electron, corrugated waveguides and associated components for, experimental work	271
Accelerators, linear electron, theoretical design	246
Accelerators, linear, travelling wave	255
Alnico, ΔE effect in (Letter)	391
Angular distribution in internal pair creation (Corrigendum)	296
Aspheric surfaces, design	494
Atmosphere, propagation of radio waves in, variational method for calculation for arbitrary refractive index profile	48
Atmosphere, radio shadow effects in, produced by inversions	59
Atmospheric absorption of millimetre waves	562
Atomic heats of palladium, sodium and mercury at low temperatures	1
Auto-oscillation, of vibrating systems	393
Bagguley, D. M. S., Bleaney, B., Griffiths, J. H. E., Penrose, R. P., and Plumpton, B. I.: Paramagnetic resonance in salts of the iron group—a preliminary survey: I. Theoretical discussion	542
Bagguley, D. M. S., Bleaney, B., Griffiths, J. H. E., Penrose, R. P., and Plumpton, B. I.: Paramagnetic resonance in salts of the iron group—a preliminary survey: II. Experimental results	551
Barrow, R. F., and Mulcahy, F. R.: The B-N band-system of silver iodide (Letter)	99
Beckman, L.: The influence of a transverse magnetic field on a cylindrical plasma	515
Benoît's method of exact fractions, modification of	194
Beta-decay, dependence on sign of change in nuclear spin (Letter)	481
Beta-disintegration, and the Sargent diagram	452
Beta-particles, fast, scattering through large angles by N nuclei	532
Beta-rays, from ^{35}S in aluminium, absorption coefficient (Letter)	485
Beta-spectrography, permanent magnet double β -ray spectrograph with coincidence counting: I. Experiments with thorium-active deposit	466
Betatrions, effect of azimuthal inhomogeneities in magnetic field	284
Bleaney, B., <i>see</i> Bagguley, D.M. S.	
B-N band system of silver iodide (Letter)	99
Bolton, H. C.: The measurement of dielectric constant in a waveguide (Letter)	294
Borysowski, Z., <i>see</i> Gurney, C.	
Bryant, A. R.: The pressures exerted by an underwater explosion bubble on a circular target plate in a disc-like baffle	341
Bunyan, D. E., Lundby, A., Ward, A. H., and Walker, D.: The delayed coincidence method in the study of radioactivity, with application to isomerism in ^{181}Ta	300
Butler, C. C., <i>see</i> Rochester, G. D.	
Byard, S.: Note on the impedance variations of an electro-acoustic transducer in a reflecting field	478
Caffyn, J. E.: A correction in viscometry due to varying hydrostatic head	367
Cathodes, oxide coated, emission of negative ions from	217
Cathodes, oxide coated, ionic conductivity	521
Chackett, K. F., and Martin, G. R.: An isotopic abundance rule (Letter)	197
Champion, F. C., and Roy, R. R.: The scattering of fast β -particles through large angles by nitrogen nuclei	532
Chapman, S.: <i>Appendix to</i> The radial variation of the earth's magnetic field	373
Chapman, S.: The dipole moment of the supposed fundamental magnetic field rotation (Letter)	95
Charged particles, observed in photographic plate, mass determination of	183
Chaudhri, R. M., and Kahn, A. W.: Emission of secondary electrons from nickel and molybdenum by neutral atoms of mercury and potassium	526

Chow, C. S. : The thermal conductivity of some insulating materials at low temperatures	206.
C.I.E. trichromatic system, distribution coefficients for calculation of colours in (Letter)	198.
Collie, C. H., and Grace, M. A. : The absorption coefficient of the β -rays from ^{34}S in aluminium (Letter)	485.
Colours in C.I.E. trichromatic system, distribution coefficients for radiator of 2450°K . (Letter)	198.
Coma image, elementary, diffraction structure of	147.
Conductivity, thermal, of insulating materials at low temperatures	206.
Cooper, R. I. B. : The attenuation of ultra-high-frequency electromagnetic radiation by rocks (Letter)	482.
Cooper, R. I. B. : The attenuation of ultra-high-frequency electromagnetic radiation by rocks	40.
Cosmic-ray intensity and height of meson formation, temperature effect	34.
Cosmic-ray showers, penetrating particles in : I. Heavily ionizing particles	307.
Cosmic-ray showers, penetrating particles in : II. Lightly ionizing particles	535.
Coulson, C. A., Craig, D. P., and Maccoll, A. : Electronic structure of fulvene	22.
Counters, Geiger, analysis of spurious counts	312.
Craggs, J. D., and Meek, J. M. : The initiation of low pressure glow discharges	327.
Craig, D. P., <i>see</i> Coulson, C. A.	
Creep, Andrade equation, method of fitting to experimental results	510.
Creep, transient, in metals, theory of	201.
Crystals, growth of (Guthrie Lecture)	403.
 Davies, D. R. : Shock waves in air at very high pressures	105.
Delayed coincidence method in study of radioactivity, and isomerism in ^{181}Ta	300.
Dielectric constant, measurement in waveguide (Letter)	294.
Dielectrics, liquid, detection of heating at low frequencies (Letter)	97.
Diffraction and optical image formation	158.
Diffraction structure of elementary coma image	147.
Dingle, R. B. : The theory of the propagation of second sound in helium II	9.
Dipole moment of supposed fundamental magnetic field rotation (Letter)	95.
Distribution coefficients for calculation of colours in C.I.E. trichromatic system for radiator at 2450°K . (Letter)	198.
Dodd, C. : The detection of heating in liquid dielectrics at low frequencies (Letter)	97.
Duperier, A. : The temperature effect on cosmic-ray intensity and the height of meson formation	34.
 Earth's magnetic field, radial variation	373.
Eastwood, E., and Mercer, K. A. : A study of transient radar echoes from the ionosphere	122.
Electromagnetic radiation, U.H.F., attenuation by rocks	40.
Electromagnetic radiation, U.H.F., attenuation by rocks (Letter)	482.
Electromagnetic waves, millimetre, atmospheric absorption	562.
Electron diffraction from anthracene, with reference to the diffuse background	430.
Electron diffraction, Kikuchi line patterns, interpretation and application : II. Methods of indexing patterns	416.
Electronic structure of fulvene	22.
Elleman, A. J., and Wilman, H. : The structure and growth of PbS deposits on rock-salt substrates	164.
Emission of negative ions from oxide coated cathodes	217.
Emission of secondary electrons from Ni and Mo by neutral atoms of Hg and Pt	526.
Energy of nuclei, observations concerning	25.
Exact fractions, modification of Benoît's method	194.
Explosion bubble, underwater, pressure exerted on circular target plate	341.
Eye movement of prone observer, measurement	489.

Feather, N., Kyles, J., and Pringle, R. W. : Investigations using a permanent magnet double β -ray spectrograph with coincidence counting : I. Experiments with thorium active deposit	466
Feather, N., and Richardson, H. O. W. : β -disintegration and the Sargent diagram	452
Foster, H. G., <i>see</i> Huxley, L. G. H.	
Fracture, delayed, of materials under tension, torsion and compression	446
Freezing-in of nuclear equilibrium (Letter)	96
Fulvene, electronic structure	22
Garlick, G. F. J., and Gibson, A. F. : The electron trap mechanism of luminescence in sulphide and silicate phosphors (Discussion)	101
Gibson, A. F., <i>see</i> Garlick, G. F. J.	
Glass, mechanism of reversible thermal after-effects at ordinary temperatures	320
Glow discharges, low pressure, initiation	327
Glueckauf, E. : Some observations concerning the energy of nuclei	25
Goldschmidt-Clermont, Y., King, D. T., Muirhead, H., and Ritson, D. M. : Determination of the masses of charged particles observed in the photographic plate	183
Gorter, C. J. : Spectroscopy at low frequencies	541
Goward, F. K. : Effect of azimuthal inhomogeneities in the magnetic field of a betatron or synchrotron	284
Goward, F. K., and Wilkins, J. J. : Validity of two-dimensional design of synchrotron pole-faces (Letter)	580
Grace, M. A., <i>see</i> Collic, C. H.	
Griffiths, J. H. E., <i>see</i> Bagguley, D. M. S.	
Gurney, C., and Boryowski, Z. : Delayed fracture of materials under tension, torsion and compression	446
Guthrie Lecture : The growth of crystals	403
Heating, detection of, in liquid dielectrics at low frequencies (Letter)	97
Hird, S. A. E., and Stephenson, H. F. : Distribution coefficients for the calculation of colours in the C.I.E. trichromatic system for a total radiator at 2450° K. (Letter)	198
Horton, G. K. : Angular distribution in internal pair creation (Corrigendum)	296
Huxley, L. G. H., Foster, H. G., and Newton, C. C. : Measurements of the interaction of radio waves in the ionosphere	134
Impact of solid on a liquid surface	352
Impedance variations of electro-acoustic transducer in reflecting field	478
Infra-red response of lead telluride photoconductors, effect of room temperature radiation (Letter)	486
Insulating materials, thermal conductivity at low temperatures	206
Interaction of radio waves in ionosphere, measurements of	134
Interferometer, Hilger prism, interference path of two neighbouring radiations	504
Inversions, radio shadow effects in atmosphere produced by	59
Ionic conductivity of oxide coated cathodes	521
Ionosphere, curved, equivalent path and absorption for oblique incidence on	78
Ionosphere, interaction of radio waves in	134
Ionosphere, transient radar echoes from	122
Ion source, radio-frequency, performance of new (Letter)	483
Isotopic abundance rule (Letter)	197
Isotopic number, and nuclear photo-effect	579
Jaeger, J. C. : Equivalent path and absorption for oblique incidence on a curved ionospheric region	78
Jensen, P. : Nuclear photo-effect and isotopic number (Letter)	579
Jones, G. O. : The mechanism of the reversible thermal after-effects in glass at ordinary temperatures	320

Index

597

PAGE

Kahn, A. W., <i>see</i> Chaudhri, R. M.	
Kennedy, A. J. : Methods of fitting Andrade's creep equation to experimental results	510
Kikuchi-line patterns, interpretation and application : II. Methods of indexing patterns	416
King, D. T., <i>see</i> Goldschmidt-Clermont, Y.	
Kingslake, R. : The diffraction structure of the elementary coma image	147
Kyles, J., <i>see</i> Feather, N.	
Lamont, H. R. L. : Atmospheric absorption of millimetre waves	562
Lattes, C. M. G., Occhialini, G. P. S., and Powell, C. F. : A determination of the ratio of the masses of π - and μ -mesons by the method of grain-counting	173
Lead sulphide deposits on rocksalt substrates, structure and growth	164
Linear accelerators, <i>see</i> Accelerators, linear.	
Liquid surface, impact of a solid on	352
Loach, B. G., <i>see</i> Mullett, L. B.	
Lord, M. P. : The measurement of the eye movement of a prone observer during monocular fixation	489
Luminescence, electron trap mechanism in sulphide and silicate phosphors (Discussion)	101
Lundby, A., <i>see</i> Bunyan, D. E.	
Maccoll, A., <i>see</i> Coulson, C. A.	
Macfarlane, G. G. : The application of a variational method to the calculation of radio wave propagation curves for an arbitrary refractive index profile in the atmosphere	48
Mackay, D. M. : Some new aspects of high-speed oscillography	235
McPetrie, J. S., and Saxton, J. A. : The attenuation of ultra-high-frequency electromagnetic radiation by rocks (Letter)	482
Magnetic field rotation, fundamental, dipole moment of (Letter)	95
Magnetic field, transverse, influence on cylindrical plasma	515
Magnetic resonance, nuclear	569
Martin, G. R., <i>see</i> Chackett, K. F.	
Mass of charged particles observed in photographic plate, determination of	183
Masses of π - and μ -mesons, ratio determination by method of grain-counting	173
Meek, J. M., <i>see</i> Craggs, J. D.	
Mercer, K. A., <i>see</i> Eastwood, E.	
Meson formation, height of, and cosmic-ray intensity, temperature effect	34
Mesons, π - and μ -, mass ratio determination by method of grain-counting	173
Metals, theory of transient creep in	201
Michaelis, E. G. : A modification of Benoît's method of exact fractions	194
Moffatt, J., <i>see</i> Thonemann, P. C.	
Muirhead, H., <i>see</i> Goldschmidt-Clermont, Y.	
Mulcahy, F. R., <i>see</i> Barrow, R. F.	
Mullett, L. B., and Loach, B. G. : Experimental work on corrugated waveguides and associated components for linear electron accelerators	271
Newton, C. C., <i>see</i> Huxley, L. G. H.	
Nuclear equilibrium, freezing-in of (Letter)	96
Nuclear forces, Schwinger's theory	297
Nuclear magnetic resonance	569
Nuclear spin, dependence of beta-decay on sign of change (Letter)	481
Nuclei, observations concerning energy of	25
Occhialini, G. P. S., <i>see</i> Lattes, C. M. G.	
Oscillography, high speed, new aspects	235

	PAGE
Paramagnetic resonance in salts of iron group : I. Theory	542
Paramagnetic resonance in salts of iron group : II. Experiment	551
Paterson, S. : The reflection of a plane shock wave at a gaseous interface	119
Penetrating particles in cosmic-ray showers : I. Heavily ionizing particles	307
Penetrating particles in cosmic-ray showers : II. Lightly ionizing particles	535
Penrose, R. P., <i>see</i> Bagguley, D. M. S.	
Phosphors, electron trap mechanism of luminescence in (Discussion)	101
Photoconductors, lead telluride, infra-red response, effect of room temperature radiation on (Letter)	486
Photo-effect, nuclear, and isotopic number (Letter)	579
Pickard, G. L., and Simon, F. E. : The atomic heats of palladium, sodium and mercury at low temperatures	1
Plasma, cylindrical, influence of transverse magnetic field on	515
Plumpton, B. I., <i>see</i> Bagguley, D. M. S.	
Powell, C. F., <i>see</i> Lattes, C. M. G.	
Pressure exerted by under-water explosion bubble on circular target plate	341
Price, W. L. : Radio shadow effects produced in the atmosphere by inversions	59
Pringle, R. W., <i>see</i> Feather, N.	
Propagation, radio wave, in atmosphere, for arbitrary refractive index profile	48
Propagation of second sound in helium II	9
Putman, J. L. : Analysis of spurious counts in Geiger counters	312
Radar echoes, transient, from ionosphere	122
Radioactivity, delayed coincidence method in study, and isomerism in ^{181}Ta	300
Radio shadow effects in atmosphere produced by inversions	59
Ramsey, W. H. : On Schwinger's theory of nuclear forces	297
Reflectivity, metallic, variation with temperature of (Discussion)	101
Richardson, E. G. : The impact of a solid on a liquid surface	352
Richardson, H. O. W., <i>see</i> Feather, N.	
Ritson, D. M., <i>see</i> Goldschmidt-Clermont, Y.	
Roaf, D., <i>see</i> Thonemann, P. C.	
Rocard, Y. : Sur les conditions d'auto-oscillation des systèmes vibrants	393
Rochester, G. D., and Butler, C. C. : The penetrating particles in cosmic-ray showers : I. Heavily ionizing particles	307
Rochester, G. D., and Butler, C. C. : The penetrating particles in cosmic-ray showers : II. The lightly ionizing particles	535
Roy, R. R., <i>see</i> Champion, F. C.	
Runcorn, S. K. : The radial variation of the earth's magnetic field	373
Sanders, J. H., <i>see</i> Thonemann, P. C.	
Sargent diagram, beta-disintegration and	452
Saxton, J. A., <i>see</i> McPetrie, J. S.	
Scattering of fast β -particles through large angles by N nuclei	532
Schneider, E. E. : Note on nuclear magnetic resonance	569
Schwinger's theory of nuclear forces	297
Secondary electrons, emission from Ni and Mo by neutral atoms of Hg and Pt	526
Second sound, propagation in helium II	9
Shersby-Harvie, R. B. R. : Travelling wave linear accelerators	255
Shock waves in air at very high pressures	105
Shock waves, plane, reflection at gaseous interface	119
Silver iodide, B-N band system of (Letter)	99
Simon, F. E., <i>see</i> Pickard, G. L.	
Simpson, O. : The effect of room temperature radiation on the infra-red response of lead telluride photoconductors (Letter)	486
Skrebowski, J. : Electron diffraction from anthracene with special reference to the diffuse background	430
Sloane, R. H., and Watt, C. S. : The emission of negative ions from oxide coated cathodes	217
Smith, C. L. : A theory of transient creep in metals	201

Index

599

PAGE

Spectroscopy at low frequencies	541
Stephenson, H. F., <i>see</i> Hird, S. A. E.	
Stokes, A. R. : A numerical Fourier-analysis method for the correction of widths and shapes of lines on x-ray powder photographs	382
Strachan, C. : Dependence of β -decay on the sign of the change in nuclear spin (Letter)	481
Street, R., and Woolley, J. C. : A note on the ΔE effect in alnico (Letter)	391
Synchrotron pole-faces, validity of two-dimensional design (Letter)	580
Synchrotrons, effect of azimuthal inhomogeneities in magnetic field	284
Taylor, A. : On a new type of rotating anode x-ray tube	86
Temperature effect on cosmic-ray intensity and height of meson formation	34
Thermal after-effects, reversible mechanism in glass at ordinary temperatures	320
Thomson, G. P. : The growth of crystals (Thirty-second Guthrie Lecture)	403
Thonemann, P. C., Moffatt, J., Roaf, D., and Sanders, J. H. : The performance of a new radio-frequency ion source	482
Thorium-active deposit, experiments with permanent magnet double β -ray spectrograph	466
Transducers, electro-acoustic, impedance variations in reflecting field	478
Travelling wave linear accelerators	255
Tsien, L. C. : Interference pattern of two neighbouring radiations on a Hilger prism interferometer	504
Ubbelohde, A. R. : The freezing-in of nuclear equilibrium (Letter)	96
Ultra-high-frequency electromagnetic radiation, attenuation by rocks	40
Ultra-high-frequency electromagnetic radiation, attenuation by rocks (Letter)	482
Vibrating systems, auto-oscillation of	393
Viscometry, correction, due to varying hydrostatic head	367
Wagener, S. : Note on the ionic conductivity of oxide coated cathodes	521
Walker, D., <i>see</i> Bunyan, D. E.	
Walkinshaw, W. : Theoretical design of linear accelerator for electrons	246
Ward, A. H., <i>see</i> Bunyan, D. E.	
Watt, C. S., <i>see</i> Sloane, R. H.	
Waveguides, corrugated, and other components, for linear electron accelerators, experimental work on	271
Waveguides, measurement of dielectric constant in (Letter).	294
Weil, R. : The variation with temperature of metallic reflectivity (Discussion)	101
Wilkins, J. J. :, <i>see</i> Goward, F. K.	
Wilman, H. : The interpretation and application of electron-diffraction Kikuchi-line patterns : II. Methods of indexing the patterns	416
Wilman, H., <i>see</i> Elleman, A. J.	
Wolf, E. : On the designing of aspheric surfaces	494
Woolley, J. C., <i>see</i> Street, R.	
X-ray powder photographs, line width and shape correction, Fourier-analysis method	382
X-ray tube, rotating anode, new type	86
Young's modulus, ΔE effect in alnico (Letter)	391
Zernike, F. : Diffraction and optical image formation	158

INDEX TO REVIEWS OF BOOKS

	PAGE
Andrade, E. N. da C. : <i>The Atom and its Energy</i>	583
Brailsford, F. : <i>Magnetic Materials</i>	104
Jánossy, L. <i>Cosmic Rays and Nuclear Physics</i>	584
Jauncey, G. E. M. : <i>Modern Physics</i>	200
Johnson, Martin : <i>Science and the Meanings of Truth</i>	102
Mercier, P. and A. : <i>Précis de Physique Générale, II : La Chaleur</i>	104
Millikan, Robert A. : <i>Electrons (+ and -), Protons, Photons, Neutrons, Mesotrons and Cosmic Rays</i>	583
Roberts, Arthur : <i>Radar Beacons</i>	488
Royal Society Mathematical Tables Committee : <i>Mathematical Tables for Science and Industry</i>	296
Smith, R. A. : <i>Radio Aids to Navigation</i>	582
Townsend, Sir John : <i>Electrons in Gases</i>	294

